

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 101.03 Seconds
(Without alignments)
1052.139 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLMRRLVSL.....RASFGVNPATPTADTYLQX 957

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	6.5	1751	20 AAY13493	Truncated cellulases
2	323	6.3	782	12 AAR15625	Cellulase AE-1. A
3	301.5	5.9	1749	22 AAB72314	Glutamic acid rich
4	276	5.4	1426	20 AAY13492	Truncated cellulases
5	274	5.3	406	22 ABG27250	Novel human diago
6	261.5	5.1	531	18 AAW15238	Scaffoldin protein
7	261.5	5.1	1853	19 AAW43108	C. thermocellum ce
8	244	4.8	1664	19 AAW43106	C. thermocellum OI
9	229.5	4.5	950	22 ABB71513	Drosophila melanog
10	227.5	4.4	455	22 AAU44346	Propionibacterium
11	225.5	4.4	551	18 AAW18790	Corrected Bacillus

12	219	4.3	1237	21 AAY81609	Streptococcus pneu
13	218	4.2	88	22 AAC002124	Human polyptide
14	216	4.2	237	22 AAB74609	Leishmania major P
15	213	4.1	1357	22 ABB70775	Drosophila melanog
16	213	4.1	2344	22 AAU37120	Staphylococcus aur
17	212.5	4.1	521	22 AAB48788	Acidothermus cellu
18	212.5	4.1	598	22 ABG14000	Novel human diago
19	212	4.1	1532	21 AAB40945	Human ORFX ORF709
20	211.5	4.1	521	22 AAB48787	Acidothermus cellu
21	210.5	4.1	521	17 AAR89927	A. cellulolyticus
22	210.5	4.1	521	22 AAB48786	Acidothermus cellu
23	210.5	4.1	562	21 AAY69508	Acidothermus cellu
24	209	4.1	918	20 AAY08640	S. aureus ClfB pro
25	208.5	4.1	206	18 AAW14574	Streptococcus pneu
26	208.5	4.1	493	20 AAY28850	Pectate lyase-link
27	208.5	4.1	493	20 AAY43218	Pectate lyase CBD
28	207.5	4.0	1291	19 AAW59912	Amino acid sequenc
29	206.5	4.0	875	19 AAW34987	Bankia gouldi endo
30	206	4.0	1805	22 ABB65262	Drosophila melanog
31	205	4.0	311	22 AAU51992	Propionibacterium
32	205	4.0	1180	22 ABB65175	Drosophila melanog
33	204.5	4.0	499	14 AAR42122	NK-1 cellulase. B
34	204.5	4.0	957	21 AAY59288	Human MUC11 polype
35	204.5	4.0	957	22 AAM24513	C900P predicted am
36	202.5	3.9	476	21 AAY54123	A mannanase-linker
37	202	3.9	731	22 AAB74209	Protein encoded by
38	200.5	3.9	700	12 AAR13227	Novel endoglucanas
39	200	3.9	183	18 AAW14570	Streptococcus pneu
40	200	3.9	2586	22 ABB66878	Drosophila melanog
41	199.5	3.9	490	20 AAW95602	Saccharothrix aust
42	199.5	3.9	616	20 AAY13494	Truncated cellulases
43	199	3.9	317	21 AAG08664	Arabidopsis thalia
44	199	3.9	333	21 AAG46507	Arabidopsis thalia
45	199	3.9	448	21 AAB53201	Macaca mulatta tha

ALIGNMENTS

RESULT 1

AAV13493
ID AAY13493 standard; Protein; 1751 AA.

XX AAY13493;

XX 30-JUL-1999 (first entry)

XX Truncated cellulases comprising amino acid sequence.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

XX cotton-containing fabric; stonewashing.

XX Unidentified.

XX OS

XX PN

XX EP921188-A2.

XX PD

XX 09-JUN-1999.

XX PF

XX 15-SEP-1998; 98EP-0810919.

XX PR

XX 19-SEP-1997; 97US-0932571.

XX XX

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX PA

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;

XX PI

XX Gibbs MD, Morgan H, Williams DP;

XX PI

XX Novel human diago

XX Scaffoldin protein

XX C. thermocellum ce

XX DR

XX C. thermocellum OI

XX Drosophila melanog

XX Propionibacterium

XX PT

XX New truncated cellulase proteins, useful in detergents and for

XX producing 'stonewashed' denim

XX PT

xx PS Claim 7; Page 37-41; 65pp; English.

CC The invention relates to a recombinant cellulase active protein free of

CC proteinases of native thermophilic and alkaliphilic origin, comprising

CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,

CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-

CC length sequences, or functional equivalents. Cel B5 extends from amino

CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino

CC acid K635 to N1426 in the sequence shown in AY13492; Cel E1 extends

CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel

CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to

CC K1751 and the stability region extends from amino acid E482 to G635 in

CC the sequence shown in AY13493; Cel E3/B5 is shown in AY13494. The new

CC enzymes are useful in laundry detergent compositions to prevent or

CC remove staining, backstaining or graying, for use on cellulosic

CC materials including cotton-containing fabrics. They are especially useful

CC for preventing redeposition of colorant during stonewashing, and for

CC processing of textiles where cellulose breakdown is required. The new

CC truncated enzymes show reduced redeposition of dye compared to using

CC non-truncated cellulase compositions.

xx SQ Sequence 1751 AA;

Query Match 6.5%; Score 334; DB 20; Length 1751;

Best Local Similarity 21.2%; Pred. No. 2e-09;

Matches 219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;

QY 18 SLLAATASFAVAALGVLPATITASPAHAA-----TTQP---YTWSNVAI 59

Db 186 svvaet-----aaSLAVASVVKERNsqkaasyIqhakdlfEadfttrsdagytatgyf 240

QY 60 GGGGFVDGIVFNEGAPGLIXVRTD-----IGGMVW-----DAANGRWIP 99

Db 241 tsggfiddlgw---aawlylatndssyltkaeelmseyangntwtqcdwdrvrygllim 297

QY 100 LLDVWGNNGYNGVSIADPINTNKVAAVGM-YTNSWDPNDGAILRSSDQG----- 152

Db 298 lakitgkel-ykgavernidhwdrtptkpmayitgw---gsIryattaafIacvy 351

QY 153 ATWQ-----IPLPKLGNMGR---GMGERLAVDPNNNILYFGAPSG 194

Db 352 adwsgcdsnkktkylnfaksgidyalgst--grsfvvgfntynpqphhrrn----- 400

QY 195 KGLWRSTDGATWSQMTNFPD-----VGYIANPTDTTGYQSDIOGVVW---VAFDKSSS 246

Db 401 -----ahsswansmkipeyhrhilygalvgpgsdssyndditdyvqnevacydnag 452

QY 247 SLGQASTIFGVADPNPNVFWSDGGATQWQVPGAPTGFIPHKGVFDPVNVHLYIAT-- 304

Db 453 ivgalak-----mvglygggepid--dfkaIetpndeliveskf 489

QY 305 SNTGGPYDSSGDVWKFVTSGTWTRISVPVSTDTAN-DYF-----GYSGLTIDRQ 354

Db 490 gnsqgp---nytevislynrwtg----pprvtdkIsfkfidlteliqagys----- 535

QY 355 HPNTIMVATQISWMPDTRIIFRSTDGGATWRIWDWTSYPNRSLRYVL-DISAEFWLTFGV 413

Db 536 -pdvkvdt-----yyiegkkgisgpyvwd---karnlyyvlvdfsgtk----- 575

QY 414 QPNPVPSPKLGW---MDAMATIDPPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM 469

Db 576 ypggevehkkaqkIasvpgypwdpnt-dpsykgltsqleknkyaaydnna----- 628

QY 470 VKGLEETAVNDLISP-----PSGAPLISALGDLGFTHADYTAVPSTIFT---SPVFT 519

Db 629 vwIepgaatstaptstptptptvtva-----tptptptptgsgt--g 674

QY 520 TGTSDVDAELNPSIIVRAGSFDSPSQPNDRHVAFTDGGKNWFGSPGVTGTTGTVAAAS 579

Db 675 tsgsvklyknnetsastgsirp-----wfk-----ivngg--sss 708

QY 580 ADGSRF---VM--APGDPGPVVYVAVGVFGNSWAASQGVPAQAIRSDRVNPKTEYALSNG 634

Db 709 vdlsvrkirywtvdgkpgsavg-----dw-----adigasvnt-fnfvkIssg 752

QY 635 TFYRSTDDGVTQFQVAAAGLSSGAVGMFHAVPQKE-GDLWLAASSGLYHSTNGSSWSA 693

Db 753 v-----sgadyy-levgf-ssgagqlq----pgkdtgdiqvirfnkndwsnyngaddsw 800

QY 694 ITGVSSAVNVGFKSAPSSSYPAVVFVVTGGVTGAYRSDDCGTTWVLINDDQHOYGNMG 753

Db 801 lqsmtn-----ygenakvtly-----vdg-----vIv-----wg 824

QY 754 QAITGDHANLRVYIGTNGRGIYVDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPS 813

Db 825 q-----epgga-----tpapstatpt 841

QY 814 PSFSSSPSSP 873

Db 842 ptptatptptptptptvatptaptaspgvsgsywtpses-----ygal 887

QY 874 KVOYKNDSPAGDNOIKPGLQVWNTGSSVDLSLTVTYVWFTRDGSGSTLVYNCDMAAIG 933

Db 888 kvwyangnlspstnvpkikienvttaavlsvrvywytidgeatqsv----- 938

QY 934 CGNIRASFGSVNPA 947

Db 939 -----svassinpa 947

RESULT 2

AA15625

ID AA15625 standard; Protein; 782 AA.

XX AC AA15625;

XX DT 17-MAR-1992 (first entry)

XX DE Cellulase AE-1.

XX KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose.

XX OS Aeromonas strain no. 212.

XX PN JP03251174-A.

XX PD 08-NOV-1991.

XX PF 28-FEB-1990; 90JP-0045465.

XX PR 28-FEB-1990; 90JP-0045465.

XX PA (OJIP) OJI PAPER KK.

XX DR WPI; 1991-373412/51.

XX DR N-PSDB; AAQ15178.

XX PT Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -

XX PS of opt. pH when carboxymethylcellulose is used as substrate.

XX PS Claim 2; Fig 3; 8pp; Japanese.

XX CC The sequence was deduced from the gene which was sequenced from

CC plasmid, pAEC 1, prep. by ligating chromosomal DNA contg. the

CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.

CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-

CC methylcellulose is the substrate. The N-terminal sequence: GHADT-

CC has been confirmed by Edman degradation. The gene can be used to

CC produce recombinant enzyme which is used for the effective utilis-

CC ation of biomass resources and the mfr. of pharmaceuticals and

CC foodstuffs, and also for the detergent and deinking of waste paper.

XX SQ Sequence 782 AA;

XX	AC	AAB72314;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
----	----	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

XX	AAB72314;				
AC					
XX	16-MAY-2001 (first entry)				
DT					
XX					
DE	Glutamic acid rich protein-like protein amino acid sequence.				
XX					
DE	Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;				
KW	NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;				
KW	pathogen resistance; abscisic acid metabolism.				
XX					
OS	Synechocystis sp.				
XX					
PN	WO200112801-A2.				
XX					
PD	22-FEB-2001.				
XX					
XX	17-AUG-2000; 2000WO-US22961.				
PF					
XX					
PR	18-AUG-1999; 99US-0149656.				
PR	23-MAY-2000; 2000US-0206405.				
XX					
XX	{PION-} PIONEER HI-BRED INT INC.				
PA	{CURA-} CURAGEN CORP.				
PI	Bidney DL, Crasta OR, Hu X, Lu G;				
XX					
DR	WPI; 2001-211215/21.				
XX					
PT	Novel isolated defence-related signalling gene isolated from sunflower				
PT	encoding neoxanthin cleavage enzyme, amino acid permease or glutamic				
PT	acid-rich protein useful for increasing resistance of plant to a				
PT	pathogen				
XX					
PS	Example 1; Page 128-135; 135pp; English.				
XX					
CC	This invention relates to defence-related signalling genes isolated from				
CC	the sunflower (Helianthus annuus). The genes encode a neoxanthin cleavage				
CC	enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich				
CC	protein (GRP). The signalling gene is useful for increasing the				
CC	resistance of a plant to a pathogen such as fungus, virus, bacterium,				
CC	nematode or insect (e.g. European corn borer), preferably				
CC	Sclerotinia spp., Phoma spp., or Phomopsis spp, by stably incorporating a				
CC	construct containing the gene into the genome of the plant. The gene is				
CC	useful for regulating gene expression in a plant, in response to a				
CC	stimulus such as infection with a pathogen, damage from a pathogen,				
CC	hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,				
CC	oxalic acid or expression of a gene encoding oxalic acid oxidase. The				
CC	genes are also useful for stem-preferred regulation of gene expression in				
CC	a plant. The genes are useful in agriculture, particularly in the				
CC	breeding of crop plants with improved agronomic traits, for modifying				
CC	abscisic acid (ABA) metabolism and for modifying amino acid transport and				
CC	content in plants. The present sequence represents a glutamic acid rich				
CC	protein (GRP)-like protein isolated from Synechocystis sp. The protein is				
CC	used in the characterisation of sunflower AAP.				
XX					
SQ	Sequence 1749 AA;				
	Query Match 5.9%; Score 301.5; DB 22; Length 1749;				
	Best Local Similarity 19.0%; Pred. No. 1.1e-07;				
	Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;				
QY	17 VSLAATAAFVAAALGVLPATASPAHAATTQPTWSNVAIGGG----- 62				
Db	23 lallacsssfsg---nvlaqnitpadgtgttvdagqnfnggsgslgdgqnlhslq 78				
QY	63 -----GFVDGIVFNEGAGPILYVRTDIGMVRWDA 92				
Db	79 qfgldgqianflsnpdilnrltrivggdasiingliqvsngnanflmnpagmifgpn 138				
QY	93 A-----NGRWIPLLDWGVNMGVGVVS----- 116				

Db 139 sinvpdgvvttsaigfngdqv-----fsdndynalnignpsqfardlanpplina 193
 QY 117 -----IAADPINTKGVAAAGVMTNSWDPNDGAILRSS----- 149
 Db 194 gdlsvtegnlftlagnivntgslaaaggnitvaavpgqnririsaqgallsleverspq 253
 QY 150 -DQATWQITPLPFKLGNNPGRMGRLAVDPND-----NIFYGAPSGKGLWRSTD 202
 Db 254 mngggsfsvldlptlltqgaenldlg--lavqngsvttingtnalvspipgsvtisgnvd 311
 QY 203 SGATWQMTNFPDVGYYIANPTTTQYS---DIQ-----GVVWAFD-KSSSSLCQASK 253
 Db 312 as---gkstniss-ggqvalagqiaavqatvdvsgnggggtvrigdggqitlplnasq 367
 QY 254 TIFGVGADPNPV-----FWSRDCGA--TWQAVPCATGFPHKGVFDPVNHVLIATSNT 307
 Db 368 tli-----dsnvvkadalltnggtvivwaddstrfsgni-----saq 406
 QY 308 GGPYDSSGDVWKFVTSCTWTRISVPSPDTANDYFGSLTIDRQHPNTIMVATQISW 367
 Db 407 ggUmgnng---fvetsg-----akslmvdtdtarvntfatmgelgt 444
 QY 368 W---POTIIFRSDDGATWTRWD---WTSYPNRSRLRVLDISAEPLWTFGVQPNPPVP 420
 Db 445 wlldpleilvgttdclladpklivsvlittslndngnvlqadqdsiaavqanfsadpsap-- 502
 QY 421 SPKLGWDEAMAIIDPFNSRMLYGTGATLYA-----TNDLTKWDSGGQIHI-- 466
 Db 503 -gntfdesptitidalfs---lgtssilfantgptintgntlvtsptfndfknkqlna 557
 QY 467 -----APMWKGLEETAVN---DLISPPS-----GAPLISALGDLG----- 498
 Db 558 nttftaggydiylfrksvnggfdllgnanfvfdggagittplksfgvtateiyvgndivt 617
 QY 499 -----GTHADVT-AVPSPTFTSPVFTTG-----TSVDYAEALNP 531
 Db 618 qnqifdgvfyglqnnltsagsvlfntmillinglqvtaqniavqpsssisaveias 677
 QY 532 SIIVRAGSDPPSQPNDRHVAF---STDGKKNWFQSEPGVTTG-----GTVA 577
 Db 678 dvlilnag-----qnvsgfnintrgnvdiqal--gnistgsivtspfgnagnvi 725
 QY 578 ASADG---SRFWAPGDPQPVVAVFGNSWA-----ASQGV 612
 Db 726 lnaggtlttgyletsdngdvtstssgntstayidtrfgdgleidslggavsiesskd 785
 QY 613 PANAQIRSDRVNPKTYVALSNGTFYRSTGVTTFQVAAAGLPSGAGVGMFHAVPCKE-- 670
 Db 786 ittafidtaysiesfnegtggnvftadgslitnyiftagkngg--diffdagesieli 843
 QY 671 -----GDLWLA-----SSGLYHSTNGGS-----SWSAITGV 697
 Db 844 dyIntygsqtsgdyveapldisgysiytggggeggnvflqaggdittsyidtsaang 903
 QY 698 SSANVVG-----FGKSAPGSSYPVAVVVG 722
 Db 904 difiqsggdtdevgylftkyegrgdvyvetgryfraidgfillgeegpftsvytagitvg- 962
 QY 723 IGGVTGAYRSDGCTWVLINDDOHOYGNWQAITGDHANLRVYICTNGR--GIYVGD 780
 Db 963 -----gsyvi-----gfggsepfllgnpi-----tngtgaissgd 994
 QY 781 GGAPSGSP-----SPSVSPSASPLSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSP 825
 Db 995 ntpigtptpiftdltnitteppepepepepepepepepepepepepepepepepepepe 1054
 QY 826 PSP 868
 Db 1055 pe 1097

RESULT 4
 AAY13492

AA13492 standard; Protein; 1426 AA.
 AAY13492;
 30-JUL-1999 (first entry)
 Truncated cellulases Cel B4/5 and Cel B5.
 Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 cotton-containing fabric; stonewashing.
 Unidentified.
 EP921188-A2.
 09-JUN-1999.
 15-SEP-1998; 98EP-0810919.
 19-SEP-1997; 97US-0932571.
 (CLRN) CLARIANT FINANCE BVI LTD.
 Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 Gibbs MD, Morgan H, Williams DP;
 WPI: 1999-315403/27.
 N-PSDB; AAX55661.
 New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim
 Claim 7; Page 34-37; 65pp; English.
 The invention relates to a recombinant cellulase active protein free of
 proteinases of native thermophilic and alkalophilic origin, comprising
 the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 length sequences, or functional equivalents, Cel B5 extends from amino
 acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 acid K335 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 K1751 and the stability region extends from amino acid E482 to G635 in
 the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 enzymes are useful in laundry detergent compositions to prevent or
 remove staining, backstaining or graying, for use on cellulosic
 materials including cotton-containing fabrics. They are especially useful
 for preventing redeposition of colorant during stonewashing, and for
 processing of textiles where cellulose breakdown is required. The new
 truncated enzymes show reduced redeposition of dye compared to using
 non-truncated cellulase compositions.

Sequence 1426 AA;

Query Match 5.4%; Score 276; DB 20; Length 1426;
 Best Local Similarity 42.3%; Pred. No. 1.8e-06;
 Matches 52; Conservative 29; Mismatches 40; Indels 2; Gaps 2;
 QY 835 SRSP 893
 Db 375 saspptvtatptptptvtatptptgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 434
 QY 894 QVVNTGSSVDLSTVTVRVWTFTRDGSSTLVVNCDAATCGGNIRASFGSVNPATPATDT 953
 Db 435 kinvngssvdlrsrkirywtvtdgdkpdsav-cdwagiasnvtfnfklsgsvgady 493
 QY 954 YLQ 956
 Db 494 yle 496

[illegible]


```
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 41331.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL15616.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 41331; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 950 AA;
SQ
Query Match 4.5%; Score 229.5; DB 22; Length 950;
Best Local Similarity 25.4%; Pred. No. 0.00033;
Matches 144; Conservative 63; Mismatches 192; Indels 169; Gaps 29;
QY 405 AEPWLTGVPQNPVPV-----PSPKLGWMDMAIDPF-----NSDRMLYGT--- 445
DB 461 appssygappqavssylpaaarpkpsysg----apsvsfvlpapsapstnygapsk 516
QY 446 ----GATLYATNDLTKWDSGGQIHIAPMVK--GLEETAVNDLISPPS---GAPLISALGD 496
DB 517 tqslgssgsgppssysya----pvappsssygapsssfqp-ispsssygapsssgss 571
QY 497 LGGFTHADVTAVPSTIFTSPVFTGTGSDVYAE LNPSIIVRAGSFDPSQNDHRVAFSTD 556
DB 572 sgfsf-----saapslysap--skges-----ggsf--qsapsssysapsas 609
QY 557 GGRNWFSGSPGGVTGTGTTAAADGSRFVWAPDQGPVVYAVGFNSWAA-----SQG 611
DB 610 ansg--gyspsapssysapssssss-----ggp-----yasapssysapssnsng 655
QY 612 VPANAQIRSDRVAPKTFYALNSGTFTYRSTDGTGVTFTQPVAAAGLPSSGAVGMFHAVFCKEG 671
DB 656 gypaa-----psssysapsas---ansgg-----syapsaps 684
QY 672 DLWLAASSGLYHSTNGG-----SSWSAITGVSSAVNWGFGKSAPGSSYPVAVFVGTIG 724
```

```
DB 685 ssysapspg---snsggypaapsssysaps---psansggypasapsssysapsnsng 739
QY 725 GVTGAYRSDDCGTTWVLINDDQHGYGNWGOAITGDHANLRRVYIGTNGRGIVVGDIGGAP 784
DB 740 gpyaaapsss-----ysapsssssggypypsap 767
QY 785 SGS---PSPSVS-----PSASPSLSPPSPSS--SPSPSPSPSSSPSSSPSPSPS--PSP 834
DB 768 ssysapsesslsaggypaapsssysaapspssnsggypaapspnsysapiappsssygap 827
QY 835 SRSPSPASPPSPSSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 894
DB 828 asgppsfsf-apsssygapstgsgsssfsgsssfagassssaagypsapsssygap--- 883
QY 895 VVNTGSSVDLSTVTVYWFTRDGGST 922
DB 884 --stgshfsaapsssysaappaggss 909
RESULT 10
AAU44346
ID AAU44346 standard; Protein; 455 AA.
XX
AC AAU44346;
XX
DT 27-FEB-2002 (first entry)
XX
DE DT
XX
XX Propionibacterium acnes immunogenic protein #5242.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
PR 02-JUN-2000; 2000US-208841P.
XX
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX
DR N-PSDB; AAS59522.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
PS Claim 4; SEQ ID NO 5341; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 39.87 Seconds
(without alignments)
586.288 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	6.5	1751	4	US-09-136-574A-44
2	276	5.4	1426	4	US-09-136-574A-43
3	225.5	4.4	551	2	US-09-033-537A-1
4	210.5	4.1	521	1	US-08-276-213-3
5	208.5	4.1	493	4	US-09-198-956-10
6	208.5	4.1	493	4	US-09-198-955A-12
7	200.5	3.9	700	2	US-07-862-588B-2
8	199.5	3.9	490	4	US-09-109-841-2
9	199.5	3.9	616	4	US-09-136-574A-47
10	191.5	3.7	412	1	US-08-313-288B-18
11	189	3.7	933	3	US-08-293-728-2
12	189	3.7	933	3	US-09-421-868-2
13	186	3.6	423	2	US-08-760-797A-1
14	185.5	3.6	424	4	US-08-932-929B-1
15	185.5	3.6	1719	2	US-08-459-568-4
16	185.5	3.6	1719	2	US-08-399-411-4
17	185.5	3.6	1719	3	US-08-516-859A-4
18	185	3.6	334	6	5202236-3
19	184	3.6	331	6	5202236-37
20	181	3.5	167	5	PCT-US95-13813-9
21	180.5	3.5	424	2	US-08-760-797A-3
22	180.5	3.5	424	4	US-08-932-929B-3
23	179.5	3.5	1481	2	US-08-616-844-40
24	179.5	3.5	1481	2	US-08-599-654-40
25	179.5	3.5	1481	3	US-08-944-868A-40
26	179.5	3.5	1481	3	US-08-944-423A-40
27	179.5	3.5	1481	3	US-08-944-496-40

28	178	3.5	214	1	US-08-217-327-4
29	176.5	3.4	254	1	US-07-667-276A-6
30	175	3.4	126	1	US-08-395-602A-4
31	175	3.4	126	2	US-08-021-625D-4
32	175	3.4	531	2	US-07-862-588B-7
33	175	3.4	1581	4	US-09-110-517-2
34	171	3.3	303	4	US-08-818-112-92
35	171	3.3	303	4	US-08-818-111-93
36	171	3.3	303	4	US-09-056-556-92
37	171	3.3	786	4	US-09-103-429A-3
38	168.5	3.3	162	3	US-08-983-045-2
39	168.5	3.3	560	3	US-08-983-045-4
40	168.5	3.3	772	2	US-08-410-784A-5
41	167	3.3	160	6	5219987-7
42	167	3.3	163	5	PCT-US93-08435-2
43	167	3.3	164	5	PCT-US93-08435-1
44	167	3.3	1064	1	US-08-642-255-62
45	166.5	3.2	805	4	US-09-103-429A-4

ALIGNMENTS

RESULT 1
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 6.5%; Score 334; DB 4; Length 1751;
Best Local Similarity 21.2%; Pred. No. 1e-13;
Matches 219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;

QY 18 SLLAATASFAVAALGALPLAITASPAHAA-----TTQP---YTWNSVAI 59
DB 186 SVAET-----AASLAVASVVIKERNOKAASYLOHAKDLFEFADITRSAGTAATGYF 240
QY 60 GGGGFDGIVNEGAPGILYRTD-----ICGMYRW-----DAANGRWIP 99
DB 241 TSGGFIDDLGW---AAVWLVIATNDSSYLTKAEELMSEYANGTWTQCDWDRYGLIM 297
QY 100 LLDWVGNHNGYGVSIADPINTNKVAAVGM-YTNSNDPNDGAILRSSDQ-----152
DB 298 LAKITGKEL---YKGAVERNLDHWDRTITPKGMAYLTGW---GSLRYATTAAFLACVY 351
QY 153 ATWQ-----ITPLPFKLGNNMGR-----GMGERLAVDPNNDNILYFGAPSG 194
DB 352 ADWSGCDNSKTKYLNAFAKQIDYALGST--GRSFVVGFGTNYPOHPHNR-----400
QY 195 KGLWRSDSGATWSQMTNFPD-----VGYIANPTDTTGYQSDIQGVW--VAFDKSS 246
DB 401 -----AHSSWANSKMKIPEYHRHILYGALVGGPGSDSDYNDIDITDYQNEVACDYNAG 452
QY 247 SLGQASKTIFGVADPNPNFWSRDGGATWQAVPGAPTGFPHKGVDPVNHVLYIAT--304
DB 453 IVGALAK-----MYOLYGGEPID--DFKAIETPTNDEIFVESKF 489
QY 305 SNTGGPYDGGSDVWKFVSVTSGTWTRISVPSTDTAN-DYF-----GYSGLTIDRQ 354
DB 490 GNSGCP---NTEVISIYINRTGW---PPRVTDKLSFKYFIDLTELQAGYS-----535
QY 355 HPNTIMVATOISWNPDTIIFRSDGGATWTRINDWTSYPNRSRYVL-DISAPWLTFGV 413
DB 536 -PDVVKYDT-----YIEGGKISGYPVWD---KRNRIYVVLVDFSGTK-----I 575
QY 414 QNPVPVSPKLGW---MDEAMADPNSDRMLYGTGATLYATNDLTWKDSGQIHIAPM 459
DB 576 YPGGEVEHKKQAFKISVPOGYWDPDIN-DPSYKGLTSQLEKNKYIAAYDNN-----L 628
QY 470 VKGLEATAVNDLISP-----PSGAPLISALGDLGGFTHADYAVPSTFT---SPVFT 519
DB 629 VWGLEPGAATSTAPTSTPTPTPTPTVTA-----TPTPTPTPTPTGSP--G 674
QY 520 TGSVDVYAEINPSIIVRAGSFDPSQPNDRHVAFTDGGKNWFGSGPEPGVTTGGTVAAS 579
DB 675 TGSVKVLYKNETSASTGSIKP-----WPK-----IVNGG---SSS 708
QY 580 ADGSRF---VW---APGDPGPVYVAVGFGNSWAASQGVAPNAQIRSDRVNPKTFYALNSG 634
DB 709 VDLRSVKIRYWTVDGDKPOSACV-----DW-----AQIGASNVT-FNFVKLSSG 752
QY 635 TFRSTDCGVTFPQVAAAGLSSGAVGVHFAVPGKE-GDLWLAAASGLYHSTNGGSSWSA 693
DB 753 V-----SGADYY-LEVGF-SSGAGQLQ-----PCKDGTQVRFNKNDWSENQADDSW 800
QY 694 ITGVSSAVNVGFGKSPGSSYPAVEVVTGIGTGYAVRSDDCGTWVLINDDQHOYCNMG 753
DB 801 LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG 824
QY 754 QAITGDHANLRVYIGTNGRGIYVIGDIGGAPSGSPSPSVSPASPSLSPPSPSSPS 813
DB 825 Q-----BPGGA-----TPAPTSTATPT 841
QY 814 PPSPPSSPSSP 873
DB 842 PTPATPT 887
QY 874 KVOYKKNDSAPGDNQIKPGLQVNTGSSVVDLSTVTVRVWFTRDGSSSTLVYNCNDAIIG 933

DB 888 KVMYANGNLSSPTNLNPKIKIENVTGTAVDLSRVKRVRYWYWTIDGATQSV-----938

QY 934 CGNIRASFGSVNPA 947

DB 939 -----SVASSINPA 947

RESULT 2

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 5.4%; Score 276; DB 4; Length 1426;
Best Local Similarity 42.3%; Pred. No. 5.2e-10;
Matches 52; Conservative 29; Mismatches 40; Indels 2; Gaps 2;
QY 835 SRSPSPASPSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 893
DB 375 SASPSPVTATPT 434
QY 894 QVYNTGSSVVDLSTVTVRVWFTRDGSSSTLVYNCNDAIIGCGNIRASFGSVNPA 953
DB 435 KIVNGSSVVDLSTVTVRVWFTRDGSSSTLVYNCNDAIIGCGNIRASFGSVNPA 493
QY 954 YLQ 956

Db 41 LWSRDYRSMDDQIKS--LGYNTIRLPYSDDILKPGTMPNSINFYQMNQDLOGLTSLQVMD 98
QY 481 LIPSPGAPLISALDGLG-----GFTHADVTAVPSIFTSPTFTTGTSDVYAEI----- 529
Db 99 KI-----VAVAGQIGLRIILDRHRPCGSGSALWYTSVSEATWISDLQALQARYKG 150
QY 530 NPSIIRAGSFDPSSOPNDRHVAFSDGGKWFQGGSEPGG-----VTTGTVAA 578
Db 151 NPTVW---GFDLHNEPHDPACWCGDPSIDWRLAERAGNAVLSVNPNNLLIFVEG--VQ 204
QY 579 SAGSRVRVAPGDPQPVVYAVGEGNSAASQ-----GVP-----ANQIRSDRVNPKTFY 629
Db 205 SYNGDSYWWG-----GNLQAGAGQPVVVLNPNRVLVYSAHDYATSVYPTQWF 250
QY 630 ALSNGFYRSTDDGVTQFQPVAAAGLPSGAGVGMFHAVPCKEGDLWLAASGLYHSTNGGS 689
Db 251 --SDPTF-----PNN-----MPGIWNKNW-----GYLFNQNIAP 277
QY 690 SWSAITG--VSSAVNVGFGKSAPGSSYPVAVVGTIGGVTGAYRSD-----DCG 736
Db 278 VWLGEFTTLQSTTDQTLWKLTVQYLRP-----TAQYGADSFQWTFWSWNPDSG 326
QY 737 TTWVLNDDHQYGNNGQAITGDHANLRVYICTNRRGIVYGDIGGAPSGSPSPSPSPA 796
Db 327 DTGGILKDD-----WQTVDTVRKDGYLAPT-----KSSIFDPVGA-----SA 362
QY 797 SPISLSP 856
Db 363 SP--SQSP 418
QY 857 SSSP 915
Db 419 ASGARCTAS-----YQVNSDWNGFT-----VTVAVNTSGSVATKTTWTSWTF- 461
QY 916 RDGSGSTLVYNCDAWAA 931
Db 462 --GGNOTITNS--WNA 473

RESULT 5

US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; PRIOR FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 4.1%; Score 208.5; DB 4; Length 493;
Best Local Similarity 23.5%; Pred. No. 3.6e-06;
Matches 103; Conservative 62; Mismatches 149; Indels 125; Gaps 19;

QY 568 GGYTTGGTVAASAD-----GSRFWAPDGPQPVVYAVGFG-NSWAASQGVPAANAQIRSDR 622

Db 63 GTTTSNTSASKIDVXKDVSNVSVGSGTKE-----LKGIGIKIWRANNIIRNLKIHE-- 116
QY 623 VNPRTFYALNSGTFFYSTDGGVTTFQPVAAGLPSSGAGVGMFHAVPCKEGDLWLAASGLY 682
Db 117 -----VAGS--DKDAIG-----IEGFSKNIWYDHNH-LY 142
QY 683 HSTNGGSSW-----SAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGVTGAYR 731
Db 143 HSLNVKDYDGLFDVYKROAEYITFSWNYVHDGKSNLMGSS-----D 185
QY 732 SDCGTTWVLND-----DOHQYGNNGQAITGDHANLR-----RVYIGTNGR 773
Db 186 SDNYNRTITFHHWFENLNSRVPFRFGEHGYNNFNKLIIDSGINSRMGARIENNN-- 243
QY 774 GIYVGDIGCAPSGSPSPSVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSP 826
Db 244 -----LFENAKDPIVSWYSSSPGYWHVSNKFNKFNKSMPTTSTTTYNPPYSLD 294
QY 827 SP 886
Db 295 NVDNVKSIVKQ---NAGVGKIQRRPPTPTPTSPS---ANTPVSGNLKVEFYNSNPSDTT 348
QY 887 NQIKPGLQVYNTGSSVSDLSLTVTVRYWFTFRDGSSTLVYNCDAWAI-----CGCNI 937
Db 349 NSINPOFKVTNTGSSAIDLSKLTLYRYVYVDGQKQDTFW-CDHAAIIGNSNGSYNGITSN 407
QY 938 RASFGSVNPTATPTADTYLO 956
Db 408 KGTFVKNSSSTNNADTYLE 426

RESULT 6

US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 4.1%; Score 208.5; DB 4; Length 493;
Best Local Similarity 23.5%; Pred. No. 3.6e-06;

US-09-109-841-2

Query Match 3.9%; Score 199.5; DB 4; Length 490;
Best Local Similarity 21.0%; Pred. No. 1.4e-05;
Matches 77; Conservative 36; Mismatches 90; Indels 163; Gaps 15;

QY 568 GGVTT-GCTVAASADGSRFVAPGDPQPVVYAVGFGNSWAASQGVPPANAIQIRSDRVNPK 626
DB 27 GGVTTLSNTIAQAACGR-----VDYAV--TSQHPGFGAAVTVTNLGD---PL 70

QY 627 TFVALSNGTYRSTDGCVTFQPVAAAGLPSSGAVGVFMHVPKGEG--DLWLAASSGLYHS 684
DB 71 SSWELS-WTF-----PDGGGVQQLW-----NGV-HS 94

QY 685 TNGGS-----SWSAITGVSSAVNVGFGKSPAGSSYPAFVVGVTIGGVTGAYRSDDCGTT 738
DB 95 TSGSNVTIVKEMNSGVGTNAQVQVGFN----- 122

QY 739 WVLINDQHQYGNWGOAITDHANLRRVYIGTNGRGIVYGDIGGAPSGSPSPSVSPASP 798
DB 123 -----GSWNGA---NNAPTFTLNGTSCNAGVG-----P 149

QY 799 SLSP 858
DB 150 TTEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPT 197

QY 859 SPSPTPSSSPVSGGVKVOYKNDSPACDNOIKPCLQVWNTGSSVDLSTVTVRYWTRD 918
DB 198 TPEPTP-----EPTMPVQAGQFHVDTTNGSYRAWQAASG 232

QY 919 GSSTLV 924
DB 233 SDRDLL 238

RESULT 9
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 3.9%; Score 199.5; DB 4; Length 616;
Best Local Similarity 44.3%; Pred. No. 1.9e-05;
Matches 39; Conservative 14; Mismatches 34; Indels 1; Gaps 1;

QY 869 VSGGVKVOYKNDSPACDNOIKPCLQVWNTGSSVDLSTVTVRYWTRDGSSTLVNCD 928
DB 1 MGSQVLYVKNNTASTGSRPFWKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 59

QY 929 WAAIGCGNIRASFSGSVNPATPTADTYLQ 956
DB 60 WAOIGASNVTFNFVLSGSGVSGADYYLE 87

RESULT 10
US-08-313-288B-18
; Sequence 18, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-313-288B-18

Query Match 3.7%; Score 191.5; DB 1; Length 412;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 2, 2002, 09:11:52 ; Search time 54.67 Seconds
(without alignments)
1682.047 Million cell updates/sec
Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLMRSRLVSL.....RASFGSVNPTPTADTVLQX 957
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2436	47.4	890	2	T35237
2	2037	39.7	839	2	D97013
3	1740	33.9	856	2	T00349
4	1161	22.6	707	2	F72393
5	431.5	8.4	1779	2	T31085
6	398.5	7.8	1711	2	T31337
7	397.5	7.7	1742	2	T17120
8	372.5	7.3	473	2	S50755
9	346.5	6.7	901	2	A49227
10	340.5	6.6	1331	2	A48954
11	308.5	6.0	611	2	S76211
12	301.5	5.9	1749	2	S75138
13	293	5.7	279	2	T10361
14	291	5.7	351	2	S50754
15	290.5	5.7	2232	2	T34434
16	290	5.6	1039	2	S02711
17	288.5	5.6	1032	2	T34433
18	283.5	5.5	474	2	S15921
19	262	5.1	2468	2	A83412
20	261.5	5.1	1854	2	S36859
21	254.5	5.0	913	2	S20590
22	246	4.8	915	2	A43802
23	244	4.8	879	2	A47704
24	244	4.8	1664	2	T18262
25	243.5	4.7	552	2	T08148
26	243	4.7	288	2	T17737
27	235.5	4.6	2124	2	A28452
28	229.5	4.5	496	2	T17908
29	227.5	4.4	2271	2	F90073

30	222.5	4.3	360	2	S12850
31	222.5	4.3	3570	2	T45025
32	221	4.3	2132	1	A55182
33	219	4.3	4776	2	E95206
34	218.5	4.3	3164	1	WMBEH6
35	217	4.2	241	2	AC2284
36	216	4.2	383	2	T46707
37	216	4.2	449	2	S16748
38	215.5	4.2	13055	2	T16580
39	215	4.2	339	2	T17636
40	215	4.2	13288	2	T03099
41	214	4.2	300	2	J02220
42	214	4.2	532	2	B35621
43	211.5	4.1	1275	2	T33369
44	211	4.1	602	2	AD2067
45	210	4.1	549	2	T17525

ALIGNMENTS

RESULT 1
T35237
Probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35237
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21572
A:Accession: T35237
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-890 <SEE>
A:Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.30C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5C7.30C

Query Match	47.4%	Score 2436;	DB 2;	Length 890;	
Best Local Similarity	48.6%	Pred. No. 1.8e-102;			
Matches 475;	Conservative 129;	Mismatches 248;	Indels 126;	Gaps 19;	
QY	11	MSRRRLVSLAATASF	FAVAALGVLP	IAITASPA--HAATTPQYTWNSVAIGGGFVDGI 68	
Db	1	MRRTRLITVLLA	LAAGLLA---GSP	PAASAAEPAPRAAADS	YTKNARIDGGFVPGI 57
QY	69	VNEGAPGLIYVTRDI	GGMYRDAANGRWIPLLDWYGNWNGYGVV	STAAADPINTKMW 128	
Db	58	VFNRTKDLAYARTDI	GGAYRWQESHTWTPLLDHVGWDDWGHTGVVALASDAVDPRV 117		
QY	129	AAVGMTNSWDNDGAIL	RSSDQAGTQIPLPKLGNMPPGRGMGERLAVDPNN	DLTY 188	
Db	118	AAVGTINDWDPTNGA	VLRSADRGA	SEKADLPFKLGNMPPGRGMGERLAVDPHNDVLY 177	
QY	189	FGAPSGKGLWRSTDSGAT	WSQMTNFDPVGTIYIANPTDTTGYQSDIQGVVWAFDKSS--SS 247		
Db	178	LGAPSGHGLWRSTDAG	YTVSEVTAFPNPGNYAQDPNDTSGYASDNQGITWTFDE	STGGG 237	
QY	248	LQOASKTIFGVADPNPN	VFWSRDGGATQWAVPGATGFTPHKGVDPVNVHVIYATNS	T 307	
Db	238	AGTATRTLYGVADKENA	YRSTDAGATWELAGQPTGYLAHKGVLDAENGYLYLAYS	SDT 297	
QY	308	GGPYDSSGDVWKF	SVTSGTWTRISVP	STDTANDYFGYSGLTIDRQHPNTINVA	QISW 367
Db	298	GGPYDGGKRLRYATAT	GTWTDISPA	EADT---YYGFSGLTVDRQPGTVMATAYSSW 354	
QY	368	WPDITIFRSTDGATWTRI	WDWTSYPNRSRLRYVLDIS	APFWLTFGVQPNPPVSPKLGWM 427	
Db	355	WPDQTFIRSTDSGATWS	QAWSYTSYDPDRENYTMDVSSSPWLTWGAN	PAPPEQTPKLGWM 414	
QY	428	DEMAIDPFNSDRMLY	GTGATLYATNDLTKW--DSGGQIHIA	PMVKGLEETAVNDLISPPS 486	

Db 415 TEALEIDPFSDRMVCTGATVYGTENLTNWDDEGGTFAVPMVRGLEETAENVDLASPPS 474
QY 487 GAPLISALGDLGCTHADVTAVPSTLFTSPVFTGTSDVAELNPSLIIRAGSDFDPSSQP 546
Db 475 GAPLLSALGDVGGRFHTSLTVEFSMTYSPNFTSTSLDFAETKPDVVVRAGNLD--SGP 532
QY 547 NDRHVAFTSDGGKNNFQSGSPGGVTTGGTVAASADGSRFVWAPCDPQPVVYAVGFGNSW 606
Db 533 --HIAFSTDNGANWFGGTDPSGVSGGTVAAGADGSRFVWSP--EGAGVQVYTTGFGTSW 587
QY 607 AASQGVPAQAIRSDRVNPKFTFVALSNGTFYRSTDDGGVTQPVAA--GLPSSGAVGVNFA 665
Db 588 QASTGLPAGAIIVSDRVNPKFTFVALSNGTFYRSTDDGGVTQPVAA--GLPSSGAVGVNFA 645
QY 666 VPKEGDLWLAASS-----GLYHSTNGSGSSALITGVSSAVNVGFGKSPAGSSYPVAVFVV 720
Db 646 LPGEEDVWLAGAAGDPYGLWISTGGGTFTRLPGVDAADTVFGKAAPGASYQLFTS 705
QY 721 GTGGVTVAYRSDCCGTWYLLINDHQYGNWQQAITGDHANLRVYICTNGRGIYVGD 780
Db 706 AEIGGVGRIPRSTAGATWTRVNDAAHQMGWTCGAITGDPVYGRVYVATNGRGIYGD- 764
QY 781 GGAPSGSPSPSVSPASPSLSPSPSSSPSPSPSPSSSPSPSPSPSPSPSPSPSPSPSP 840
Db 765 -----TSDTGGTDPGPGDPPT----- 782
QY 841 SASP 900
Db 783 -----TGACEVYTYTNQHPGGFQ--ADVLINTGT 811
QY 901 S-----SVDLS-----TVTVRYWFR--DGGSTLTVNCDM-----AATCGNIRA--- 939
Db 812 SANWGSLLDMSFPGGQEVETRMNAEHTQAGTSVTRNVGNWAGVAPGASVGFGTGSRSG 871
QY 940 -----SFGSVNPATPTA 951
Db 872 TNAEPGFAVAGRACPTA 889

RESULT 2
D97013
probably secreted sialidase, several ASP-boxes and dockerin domain [Imported] - Clostrid
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: D97013
R:Nolling, J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97013
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <NUR>
A:Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:g15023820; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;
Best Local Similarity 49.5%; Pred. No. 1.5e-84;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

QY 13 SRLVSLAA-----TASFAVAALGVLPATIASPAHAAT-TQPYTWSNVAIGGGGFVD 66
Db 2 NKRIVSWAGLSIITGTF-----VTHISAANKAAASVQSGYKWDNAKIGAGGYVP 54
QY 67 GIVENEGAPCLVVRTDIGYKRWDAANGRWIPLLDWGNVNNWNGVVSIAADPINTNK 126
Db 55 AVTFNTEKDLIYARTDMGAYRWKANKNWIPTD--GFSDWTMLGCSIAIDPDIATNR 112

QY 127 VWAAGMYNTSWNDPNDGAILRSSDQATWQITPLPKLGNMFGRMGERLAYDPNNDNI 186
Db 113 VYTAAGLYTNWDQENAYILSSQDKGTWKRYQLPKFVGSGMFGRMGERLQIDPNDKI 172
QY 187 LYFGAPSGKGLWRSTSGATWSQNTFPDVGTYIANPTDTTGYQSDIQGVWVAFDKSS 246
Db 173 LYLGAASGKGLWRSTSGATWSQNTFPDVGTYIANPTDTTGYQSDIQGVWVAFDKSS 230
QY 247 SLGQASKTIFVGVAD--PNNPVFSRGGATWQAVCAPGTGFIHKGVDPDVPNHLVIATS 305
Db 231 TKGSPQTFVGAADKTNNNIYVNDGGKTSWAKQPKGYLPHHGL--ASDGLMILISYS 289
QY 306 NTGGPYDGGSDGYWKFVSSTGTRISPVSTDTANDYFYSGLTIDRQHPNTIMATQI 365
Db 290 NTCGPDGSDGQWYKNTKTGWTNITPPAVGDTKS---GFGGISVDAQPNVNVVATLN 346
QY 366 SWMPDPIIRSTDDGATWTRINDWTSYPNRSLRYVLDLSAEPWLTFG-VQNPVPVSPKL 424
Db 347 RWPDEEIVRSTDDGATWTRINDWTSYPNRSLRYVLDLSAEPWLTFG-VQNPVPVSPKL 406
QY 425 GWMDEAMADPFNSRMLYCTGATLYATNDLTWKDSSGGQIHIAPVMKGLFEETA VNDLISP 484
Db 407 GWMGDLIDPFNSRMLYCTGATLYATNDLTWKDSSGGQIHIAPVMKGLFEETA VNDLISP 466
QY 485 PSGLPISALGDLGGTTHADVTAVPSTIFTSPVFTTGTSDVAELNPSLIIRAGSDFDPSS 544
Db 467 TKGQALLSAGVDDCGFYHDDITKVPKMTTPNFSAITSTIDYAESVPNFVYRNGVNDTSK 526
QY 545 QPNDRHVAFTSDGGKNNFQ--GSEPGGVTTGGTVAASADGSRFVWAPCDPQPVVYAVGFG 603
Db 527 NQDDKCGISYDGGKNNFQ--GSEPGGVTTGGTVAASADGSRFVWAPCDPQPVVYAVGFG 585
QY 604 NSWAASQGVPAQAIRSDRVNPKFTFVALSNGTFYRSTDDGGVTQPVAA--GLPSSGAVGVN 662
Db 586 NKWTPCSGLPQGAQKVRSDRVNPKFTFVALSNGTFYRSTDDGGVTQPVAA--GLPSSGAVGVN 643
QY 663 FHAVPKEGDLWLA--ASSGLYHSTNGSGSSALITGVSSAVNVGFGKSPAGSSYPVAVFVV 721
Db 644 FKTVIGHEGDIWAGKGLWHTSDGATFTKVSVDASDTVGLGSKTKTDGYPALYMDA 703
QY 722 TGGVTVGAYRSDCCGTWYLLINDHQYGNWQQAITGDHANLRVYICTNGRGIYVGDIG 781
Db 704 TIDGTAGIFRSDDEGATWTRINDAQHGYGSDYCIITGDPNKYGRVFGTNGRGIYVGDID 763
QY 782 GAPSGSPSPSPSVS 793
Db 764 GS-QPTPTPSVT 774

RESULT 3
T00349
Avicelase III - Aspergillus aculeatus
C:Species: Aspergillus aculeatus
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C:Accession: T00349
R:Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
submitted to the EMBL Data Library, June 1998
A:Description: Avicelase III from Aspergillus aculeatus.
A:Reference number: Z14141
A:Accession: T00349
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-856 <ARA>
A:Cross-references: EMBL:AB015511; NID:d1199887; PID:d102971
C:Genetics:
A:Gene: avIII
C:Superfamily: fungal cellulose-binding domain homology
F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;
Best Local Similarity 43.5%; Pred. No. 3.5e-71;
Matches 366; Conservative 139; Mismatches 297; Indels 40; Gaps 18;

Qy	30	AAAGLVPLAIATASPAHAATTPQYTHSNVAI - GGGGFVDGIYVNEGAPGILLYVRDTDGGMY 88
Db	4	SSALLCAALGLKLIADAAASQAYTKRWNVYGGGGGTPGIVFNPSAKGAYAYARTDTCGGAY 63
Qy	89	RWDAAANGRWIPLLDWVG - --WNWNGYNGVYSIAADPTNTKNVAAVGMVYTNMSDPPNDGAI 145
Db	64	RLN - SDDTWTPLMDWVGNDTHDW - --GIDALATDPDVIDRVYVAVGMVYTNNEWDPNVGSI 119
Qy	146	LRSSDQAGATWQIITPLPKLGCNMPGRGMGERLAVDPNDNDILYFGAPSGKGLWRSTDGSA 205
Db	120	LRSTDQGGDTTETKLPLPKVGCNMPGRGMGERLAVDPNKNLSILYFGARSGHGLWKSSTDYGA 179
Qy	206	TWSQMTNFPDVGTYIANPTDTTGYSQDIQGVVWAFEDKSSSLCOASKTFYFGVADPNP 265
Db	180	TWSNVTSFTGTIVYFQDSST - YTSDPVGIANVTEDTSVSGSGSATPRIFYGVADAGKS 237
Qy	266	VFWSRDGGATWQAVPGAPT - GFTPHKGVDPDVNHLIYIATNTGGYVDSGGSDGVKFSVT 324
Db	238	VFKSEDAGAWMYSGEPQGFPLPHKGLVSPPEEKLTIYSYANGAGPYDGTNGTTHVKYNT 297
Qy	325	SGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTINVATQISWMPDTIIFRSTDGCGAT 384
Db	298	SGVWTDISP - --TSLASTYXYGGLSVLDVLPQCTLVAALCNWMPDELFIRSTDSCATWS 354
Qy	385	RWDWTSYPNRLRYVDIISAEPLWTFGVQPNP - PVPSKGLGMDEAMAIDPNNSRMLY 443
Db	355	PIWEMNGYINYYISYDINAPWIOQTSTDGFPV - --RVGMVMEALAIIDPFDSNHWLY 411
Qy	444	GTGATLATNDLTKWDSGGQIHTAPVMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHA 503
Db	412	GTGLTVYGGHDLTNWDSKHNVYKSLAVGLIEEMAVLGLITPCGPALLSAVGDDGGFYHS 471
Qy	504	DVTAVPTIITSPTVTGTSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFSTDGKNKWPQ 563
Db	472	DLDAAPNQAYHTPTGTNGIDYAGNKPNSIVRSGASD - DYPT - --LALLSSNFGSTWYA 526
Qy	564	GSEPGVYTGTTAAASDGRFVWAPGDCGPVYVAVGFGNSWAAASQGVPAANAQIRSDRV 623
Db	527	DYAASTSTGTGAVALSADGDTVLLMSSTGALYKSKQG - --TLTVASSIIPSGAVIASDKS 583
Qy	624	NPKTFYALNGTFYRSTDGVTQTPVAAGLPPSSGAVGVMFHVPVKGEGDLWLAASSGLYH 683
Db	584	DNTVYFGSAGAIYVSKNTATSFETKTYS - LGSSSTTVNAI - RAHPSIAGDVWASTDKGLWH 641
Qy	684	STNGGSSWSAI - TGVSSAVNVGFKSAPGSSYPVAVFVGTIGCVTGYARSDDCGTTWLI 742
Db	642	STDYGSTFTIGSGVTACWSEFGKASSTGYSYVVIYGFITDCAAGLFKSADGAGTWQVI 701
Qy	743	NDHQHQYGNNGQA - --ITGDHANLRVYIGNRGVIYVDITGGAPSG - -----SPSPSV 792
Db	702	SDASHGFGS - GSAVNVNGDILQTYGRVFRGHERPGHLLRQOREPAGRHGDGDDTTTTSKT 760
Qy	793	SPSASPISLPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 852
Db	761	STTVSTVTLKTTTSSASTSSSTTVKTTTSSSTTSKASSTTTTKTTTTSTTSSSGTATA 820
Qy	853	SS 854
Db	821	SA 822
RESULT 4		
F72393		
		hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
		C:Species: Thermotoga maritima
		C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-200
		C:Accession: F72393
		C:Nelson: K. F. Clouton, S. B. Givon, M. Y. Dodson, B. T. Baff

C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Methanopyrus kandleri*
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316
A:Accession: F72393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-707 <ARN>
A:Cross-references: GB:AE001712; GB:AE000512; NID:g4980799; PID:g4980799
A:Experimental source: strain MSB
C:Genetics:
A:Gene: TW0305

Query Match	22.6%	Score 1161	DB 2	Length 707
Best Local Similarity	34.6%	Pred. No. 3e-45		
Matches 267	Conservative 126	Mismatches 280	Indels 98	Gaps 27
QY 32	LGVLPIAITASPAHAATTQPTWSENVAIGGGFVDGIVENEGAPGILYVRTDIDGMYRWD	91		
DB 10	LAILGVVEGAT-----FEWKSVEINGGGFVPGFIHFPASPCLYARTDVGGLTRWD	60		
QY 92	AANGRWLPLLDWGWNNNGYGVVSIADPINTNKVAAVGMVYTNWSDPNPDGAILSRSDQ	151		
DB 61	EETKRKQLDFELRRDQSDYMGVLSVALDPSDPKRIYAMTKYTDW-AGYGAILISEDY	119		
QY 152	GATWQIITPLP---FKLGNMPGRCMGERLAVDPNDNDILFEGAPSGKGLWRSTDSGATWS	208		
DB 120	GETWTIYNLNDKYIGTKVGNEDNAGRQLQVDPNFSEVLEFMGT-TKYLGKMSDEFGNKK	178		
QY 209	QMTNFPDVGTYIANPTDITGTQSDIQGVVWAFDKSSSLQOASKTIFVGVADPNPNVFW	268		
DB 179	KVDSFPST-----SVTFVLFDKSGEKGSPTRIFVGCSEPKG-1FV	219		
QY 269	SROGGATWQAVPGAPTGFIPHKGVDFDPNVHVLXIATSNMTGGPYDGSQDVWVKFSVGTW	328		
DB 220	TEDEGTTWNVLPNLPNDLPLRGKIH--DGILYVTLNALCPNGATRGAVNKYVIADQKW	277		
QY 329	TRISVPSTDIANDYCYSGILTIDRQHPNTIMATQISWMPDITIFRSTOGGATWTRWD	388		
DB 278	YDVTPMKGD-----FGYCGIDVQE--NVVIVSLDRWTPHDEIFISLNGGETWRPLLE	328		
QY 389	WTSYPNRSRLRVLDISAEPWLITFCVQPNPPVPSPKLGWMDDEAMADIPNSDRMLYGTGAT	448		
DB 329	KANE-----DINKAPWIK-----DLNP-----HWISD-VKIDPFDMNRAITFTGYG	368		
QY 449	LYATNDLTKKDSCQIHAAPWK-----GLEATVNDLISPPSGAPLISALGDLGGFTH	502		
DB 369	VWVITYELKSEK-----MGKPVKWIENRGLLEEWIQLVPPIGERPLLSAIDAMGGRH	424		
QY 503	ADVTAPESTFTSPVFTTGSTVDVAELNPSTIIVRAGSDPSSQSDNRHRAVSTDGGKNWF	562		
DB 425	ESLDTPSSMW-KPLKWTSLGIAFYONKSFVARHYI---TYP---FLSVSEDDGINWR	477		
QY 563	Q-GSEPGVTTGG---TVAASADGRFVWAPGDPQPVVYAVFGNSWAASOGVPA---NA	616		
DB 478	EIEVPEGITDGRSLSLAVSNDGKTLVMSPAN--HEVIVSSDKGSKAKAISVPVPPEFNY	535		
QY 617	QIRSDRVNPKTFYAL--SNGTFYRSTGGVTFQPVAAGLPSGAVGVMFHA---VPCKE	670		
DB 536	FPASDPVNSKFIYFDKWKNGDFLISKDGGKSFMK-GAKLPSFDNWNWVLSYFPVLADPRE	594		
QY 671	GDWLAAAS-SGLYHSTNGGSSWSAITGVSSAVNNGYFGKSPAGSSSYPAVFVYGTIGCTGA	729		
DB 595	GDWLALQWNGLYRSKDGITFERLGNVDIAYVIGFCAPKPGTDYPAIYLNGMNGVYGI	654		
QY 730	YRSDDCGTTWVLINDDHOYGNWQQAITGDHANLRVYIQTNGRGIVYVDI	780		
DB 655	FMSYDEGKTWMRIINNDKHQFG-WTHYIMIGDMNEFRIFLGTGEGRIIVGEV	704		

RESULT
T31085
5

xylanase - *Caldicellulosiruptor* sp.

C;Species: *Caldicellulosiruptor* sp.
C;Date: 03-Sep-2000 #sequences: 1071

C:Accession: T31085
R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
submitted to the EMBL Data Library, December 1997
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.
A:Reference number: Z20972
A:Accession: T31085
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1779 <MOR>
A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1
C:Genetics:
A:Note: xynC

Query Match 8.4%; Score 431.5; DB 2; Length 1779;
Best Local Similarity 19.8%; Pred. No. 5.7e-12;
Matches 252; Conservative 162; Mismatches 382; Indels 477; Gaps 53;

QY 65 VDGIVNEGAPGILYV-----RTDIGMYRDAANGRIPLLDWGNMNGYNGVVSIAAD 120
Db 20 VGTLIHQEAKAAAYTVDFEGTDTLSFFAYGKSNIA-----VDMGN-AYNGKSSIRVS 71
QY 121 PINTNKVAAVGYMTNSWDNDGAILRSSDQATWQIT-----PLPFKL-----G 165
Db 72 --NRSSIMDGVAV-----DVRNIMNNGTTVVVSAYVKHSYOKPVAFGISAVYDDG 119
QY 166 GNPGRGMRGLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTD 225
Db 120 SGVKSILIGEVAI-PN-----YWKI-----VGKWTNINIS 150
QY 226 TTGYQSDIOGVVWVAFDKSSSLGQASKTIFGVGADPN-----NPVFWSR--DGGAT--WQ 277
Db 151 VRNLLIIVVHTIVESGVNDYND-----YIQIMDNSYLSNAVTFSSGFESGTEGWQ 201
QY 278 A----VPCAGTGPPIKGVDPVNVHLYIA--TSNTGG---PYDG--SSGDVWKFVS--- 323
Db 202 ARGSGVTVPKPSVAVYNGKYS-----LYVSGRTSNHGAQIPVDILEOGKVYKISVWVY 256
QY 324 -TSGTTRISVPSTDTAND-YFGYSLGTIDROHPNTIMVATQISWMPDTIIFRSTDGGA 381
Db 257 QNSGSTQKMSLTWQRFATDPSTSYENLYNRDVPNTWVEPSGSY-----SIPAGV 308
QY 382 TWTRINDWTSYPNRSILRY-----VLDIS--AEP----- 407
Db 309 TVSELLLYVEAQNANLAFWFDLKIYDLKSLAEPWEIPEISLIEKYDYKVGVALSYKSI 368
QY 408 -----WLTFGVQPNPPVPSPKLGW 426
Db 369 ASDTEKMWLKHNSITAGNEMKPSSELLISENNYNFSKADEFVNFATSNNAIRGHTLVW 428
QY 427 MDEA-----MAIDP----- 435
Db 429 HEQTPDFFKDNGLTSLKDLALLSLKQIYTYVGVRYKGVYANDVNVNEAIDESQNGFR 488
QY 436 -----FNSDRMLYGTGATLYA-----TNDLTWKDSGGQTHIAPMVKGLTEAVNDL 481
Db 489 RSNWYNICGPEYTEKAFIWAHEADPAKLFDYNDYNTENSKQRFIYNMKSILKE----- 542
QY 482 ISPPSGAPLISALDGLGGTHADYVAPSTIFTSPVFTT--GTSVDYAEINPILIVRAGS 539
Db 543 -----KGVP-IHGIG-LQCHINLWPSISEIENTIKLFSSIPGLEIHTELDMSFYQWSS 596
QY 540 FDPSSQPN-----RHVAF-----STDGKN-----WFO 563
Db 597 TSYSTPRDLLIKQAMRYKELDFPKKYVITVNTVWGLKDDYSWLSQNGKSDYFELLED 656
QY 564 GS-----EPGGVTITGGTV----- 580
Db 657 GNYSKYAEVSLIEPTWVPVNSLTPAPPAIQVPTPTSTPTPTPTVVSATPTTAPTASPA 716
QY 581 DGSRFVWAPCDQCPVVYVAVGFNSWAAASQGVPAQAIRS-----DRVNP 625
Db 717 GGS--YWTPE-----SYGALKVYVANGNMSSTTNVLNPKIKIENVGTAVDLSRVKV 767

QY 626 KTFYALSNGTFYRSTDGGVTFQP-----VAAGLPSSGA-----VGVMFHAVP 667
Db 768 RYWTI-DGEAGSVSVASSINPAYIDRVVVKLGANAGGADYVVEVGFKSGACVLAAGQS 826
QY 668 KREGDLWLAASSGLYHSTNGGSSWSA-----ITGVSSAVNVWGFKSKAPGSSYPVAVFV 719
Db 827 TKIIRLSIQKSSGYSNQSNDYSVRSANSYITENEKVTCYIDDLVWVGREPGRNAQIKVWA 886
QY 720 VGTPIGGVTGA---YRSDDCGTT-----WVLINDDQHOYGNWGOAITGDHANLR 764
Db 887 NGNLSSPTNVLNPKIKIENVGTITAVDLRSRVKRYWYTIDGEATQSVSVASSINPAYIDRV 946
QY 765 RYVIGTNGRIVY---GDIGGA-----PSGS----- 787
Db 947 VVKLGANAGGADYVVEVGFKSGAGVLAAGOSTKEIRLSIQKSSYSNQSNDYSVRSANSY 1006
QY 788 -----PSPSVSPSASPLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 828
Db 1007 IENEKVTVGIDDLVWVGKBPSPRGTKPAG--EVPAPTPTSTPTPTPTTPTTPTTPTTPT 1064
QY 829 SP 883
Db 1065 TVTATPT 1124
QY 884 PGDNQIKPGVLQVNTGSSSVYDLSTVTVRYWFTRDGGSSTLVYNCDAALGCGNIRASFSG 943
Db 1125 ASTGSRPWFKIVNGSSSVYDLRSVRKIRVYTVDGKPOSAN-CDNAQIGASNVTFNVK 1183
QY 944 VNPATPTADTYLQ 956
Db 1184 LTSGVSGADYILE 1196

RESULT 6

T31337

1.4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C:Species: Anaerocellum thermophilum

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

A:Accession: T31337

R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144, 457-465, 1998

A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) f

omains.

A:Reference number: Z21003; MUID:98154434

A:Accession: T31337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1711 <ZVE>

A:Cross-references: EMBL:286105; NID:e1071329; PID:e350354; PIDN:CAB06786.1

C:Genetics:

A:Gene: cela

C:Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 398.5; DB 2; Length 1711;

Best Local Similarity 22.9%; Pred. No. 1.7e-10;

Matches 197; Conservative 116; Mismatches 267; Indels 281; Gaps 40;

QY 262 PNPVFWNSRGGATWQA---VPGAPTGFIPHKGVDPVNVHLYIATSNITGGYDGS--S 315

Db 25 PN---WVRN---NWRGDSALKDQDNGLDLTGGWFDAGDHVKF---NLPMSYTGTMLS 73

QY 316 GDVWKFS---VTSGTWTRISVPSTDTANDYFCYSGLTIDROHPNTIMVATQI----- 365

Db 74 WAVIEYKDAFVKSQGLEHI--LNQIEWNDYF-----VKCHPSKYVYVYVQDGSKD 124

QY 366 SWM-----PDTIIFRSTDGGATWTRINDWTSYPNRSLRYVLDISAEPWLTFCVQP 415

Db 125 AWMGPAEVMQMERPSFKVTQSSPGSVVTE---TAASLAAASIVLK-----DR 169

QY 416 NPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLTWKDWSGQTHIAPMVKGLEE 475

Db 170 NPTKAAATYLOHAK-----LYEFAEVTKSDAG-----Y 197

Qy 476 TAVNDLISPPSG-----APLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY 526

Db 198 TAANGYNSGSGFYDELSSAAWVLYLATNDSTYLTAK-----SYVQWPKISGNTIDY 252

Qy 527 -----AELNPSIIIRAGSFPDSSQPNDRHVAFTSDG-----GKNWFQ 563

Db 253 KWAHCWDVINGAAL-----LLAKITGDIYKQIIESHLDYWTGNGERIKYTPKGLAWLD 309

Qy 564 --GSEPGGVTTGTAAASD-----GSRFVWAPGDPQPVYAVGFG-- 603

Db 310 QWGLSRVATTATFLAFVYSDWVCWGPSTKKEIYRKFEGESQIDYALGSAGRS--FVWGFTN 367

Qy 604 -----NSWAASQGVPA-----NAQIRSD-- 621

Db 368 PPKRPHRTAHSSWADSQSIPSYHRHTLYGALVGGPGSDSDSYTDDISNYVNEVACDYNA 427

Qy 622 -----RVNP-----KTFYALNSGTFFYRSTDDGVTFFOPVAAGLPSGSGAVGWYFH 664

Db 428 GFVGAALAKMYLGGNFIPIDFKAIETPTNDEFF-----VEAGINASGTNFIEIK 476

Qy 665 AVPGKEGDLMAASS-----GLYHSTNGGSSAITGVSSAVNVGFKSAP----- 710

Db 477 AIVNQSG-WPAKATDLKFRYFVLDSELIKAGYSPNQLT-LSTNYNQAKVSGPYVWDA 534

Qy 711 --GSSYPAVFVGTI--GGVTGAYRSDDCGTTWVLIINDHQYQNGWQOAITGDHANLRRV 766

Db 535 SKNTYYILVDFTGLIYPGGQDYKKE-----VQFRIAPQNVQWDSNDYSFQDIKGV 588

Qy 767 YIGT-----NGRGIVYDGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 816

Db 589 SSGSVTKKIPIYLDGVKVGDCGPTGATPTPTA-----TATPTPTPTPTPTPTPTPTPT 642

Qy 817 SSSPSSP 875

Db 643 TS--TATPTPTPTPTPTPTPTPTPTPTPTA--TPTSTPTPTS-----TPSTPTPVAGGQIKV 690

Qy 876 QYKKNDSAPGDNQKPGLOVNVNTGSSSVLDSTVTVRWFTRDGGSSILVNCDAATGCG 935

Db 691 LYANKENTSTNTIRPWLKVNTGSSSIDLSRVTVRYWTVDDGKAQSAIS-DWAQIGAS 749

Qy 936 NIRASFGSVNATPTADTYLQ 956

Db 750 NVTEKFVKLSSSVSGADYYLE 770

RESULT 7

T17120

cellulase (EC 3.2.1.1) precursor, thermoactive - Caldocellum saccharolyticum

C:Species: Caldocellum saccharolyticum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000

C:Accession: T17120; A43745

R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.

Appl. Microbiol. Biotechnol. 43, 291-296, 1995

A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophilic bacterium Caldocellum saccharolyticum

A:Reference number: Z18698; MUID:95336703

A:Accession: T17120

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1742 <TEO>

A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1

R:Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 57, 694-700, 1991

A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a cellulase from the extreme thermophilic bacterium Caldocellum saccharolyticum

A:Reference number: A43745; MUID:91247819

A:Accession: A43745

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1516-1544, 'A', 1546-1742 <LUE>

A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293

C:Genetics:

A:Gene: cella

C;Keywords: glycosidase; hydrolase

Query Match 7.7%; Score 397.5; DB 2; Length 1742;
Best Local Similarity 21.7%; Pred. No. 1.9e-10;
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

Qy 262 PNPVFRSGGGATWQA-----VPGAPTGFPHKGVFDPVNVHLY-IATSNVTGGPYDSSG 316

Db 48 PN-----WVRN-----NWRGDSALKDQDNDLDTGWFAGDHVAFNLPMSYTGTMLSWAA 100

Qy 317 DVKFSVSTGVTWRISPVSTDTANDYFGYSLGTIDRQHPNTIMVATQI-----SWW- 368

Db 101 EYKDAFVKSQLEHI--LNQIEWNDIF-----VCFESKIVYIYQVGGGKDHAWMG 151

Qy 369 -----PDTIIFRSTGGATWTRIMWTSTYPNRSRLRYVLDISARPFWLTFTGQPNPPV 419

Db 152 PAEVMQMERPSFKVTQSSGSA-----VVAETAASLAASAIVLKDRN 193

Qy 420 PSPKLHMDAMALDPNSDRMLYGTGATLYATNLDLTKWDSGGQIHIAPMVKGLEETAVN 479

Db 194 PTKAATYLOHA-----KDIYFAEVTKSDG-----YTAAN 224

Qy 480 DLISPPSG-----APLISALDGLGGTHADVTAVPSTIFTSPVFTTGTSDVY-- 526

Db 225 GYNSWSGFFYDELSSAAWVLYLATNDSTYLTAK-----SYVQWPKISGNTIDYKWAH 279

Qy 527 -----AELNPSIIIRAGSFPDSSQPNDRHVAFTSDG-----GKNW----- 561

Db 280 CWDDVHNGAAL--LLAKITDKDYKQIIESHLDYWTGNGERIKYTPKGLAWLDQWS 336

Qy 562 -----FQSEPGGVTTGTTVAASADG-SRFVWAPGDPQPVYAVGFG----- 603

Db 337 LRYATTTAFALFYSDWSGCTGKKETRYKFGESQIDYALGSGRS--FVWGFTNPKR 394

Qy 604 -----NSWAASQGVPA-----NAQIRSD----- 621

Db 395 PHRTAHSSWADSQSIPSYHRHTLYGALVGGPGSDSDSYTDDISNYVNEVACDYNAGFVG 454

Qy 622 -----RVNP-----KTFYALNSGTFFYRSTDDGVTFFOPVAAGLPSGSGAVGWYFH 668

Db 455 ALAKMYLGGNFIPIDFKAIETPTNDEFF-----VEAGINASGTNFIEIKAI-- 501

Qy 669 KEGDLWLAASSGLYHSTNGGSSWA-----ITGVSSAVNVGFKSAPGSSYPAVF 718

Db 502 -----VNNQSGWPARATNKLKFRYFVLDSELIKAGYS-----PNQL 537

Qy 719 VVGTI-----GGVTGAYRSDDCGTTWVLIIND-----DQHOYGNMQOAITGDHANLRR 765

Db 538 TLSTNYNQAKVSGPYVWDSRNIYYILVDFTGLIYPGGQDYKKEVQFRIAPQNVQ- 596

Qy 766 VYGTNGRGIVYDGGAPSGS-----PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 807

Db 597 --WDNSNDYSFQDIKGVSGSVVTKKIPIYLDREDIKVWGEPEGTS-----GVSPTPTAS 648

Qy 808 SPSPPSS 867

Db 649 VTPPTPTPTA-----TPT 700

Qy 868 P-VSGGVKVOYKKNDSAPGDNQKPGLOVNVNTGSSSVLDSTVTVRWFTRDGGSSILVYN 926

Db 701 PATSGQIKVLYANKETSTNTIRPWLKVNVNSGSSSIDLSRVTVRYWTVDDGKAQSAIS 760

Qy 927 CDWAAICGNIRASFGSVNATPTADTYLQ 956

Db 761 -DWAQIGASNVTEKFVKLSSSVSGADYYLE 789

RESULT 8

S50755

hypothetical protein Vsp-3 - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii

C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000

A;Accession: S75138
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1749 <KAN>
A;Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BAAL7052.1; PID:d101778
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

```
Query Match          5.9%; Score 301.5; DB 2; Length 1749;
Best Local Similarity 19.0%; Pred. No. 3.9e-06;
Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;

QY 17 VSLAATASFAAAALGVLPIAITASPAHAATTPYWSNVAIGG----- 62
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 23 LALLACSSFSFG----NVLAQNIIPADPGTCTTVDAGNQFNIGGSLGQGNLFSLQ 78
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 63 -----GFVDGIVFNEGAPGILYVRTDIGMYRWA 92
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 79 QFGLDQGOIANFLSNPDIRNLTIRIVGDDASIIINGLIQVSGGNANFLMNPAGMIFGPN 138
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 93 A-----NCRWIPLLDWGNWNGVNVVS----- 116
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 139 SINVPGEFVVTGSAIGFNDQMFQV-----FSDNDINALIGNPSQAFDLANPGLIINA 193
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 117 -----IAADPINTNKWAAVGMVYTNWDNDGAILRSS----- 149
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 194 GDLSVTEGKNLTFLAGNIVNTGSLAAPGNTVAAVPGNQNRIRISOAGSLLSLEVEVSPQ 253
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 150 -DQATWQITPLPKLGGNPGRGMRGLAVDPNND-----NLLYFGASGKGLWRSTD 202
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 254 MNQGSFSLVDLPLTLLQGASNLDLG--LAVQPGNSVTTNGTNALVPLPGSVTISQNV 311
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 203 SGATWSQMTNFPDVGTVIANPTDTTGYS--DIO-----GVWVAFD-KSSSSLGQASK 253
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 312 AS---GKSTNISS--GGQVATAGDIAVQATVDVSGNGGGTVRIGDFGQGLTLPNASQ 367
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 254 TIFGVGADPNPV-----FWSRDGGA--TWQAVPGAPTGFIPHKGVFDPVNVHLYIATSN 307
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 368 TLI-----DSNVKADALLTGNGGTIVWADDSRFSGNI-----SAQ 406
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 308 GGPYDSSGDWKFVSVSGTWTRISPPVSTDTANDYFGYSLGTLIDROHPNTIMVATQISW 367
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 407 GGTMGNGG---FVETSG-----AKSLMWDDTARVNTFATMGELGT 444
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 368 W---PDRITIFRSTGGATWTRIMD---WTSYPNRSRYVLDISAEPLWTFGVQNPVPV 420
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 445 WLLDPLEIIVGTTDLLADPKLVSVLTITSLDNGVILQADQSIQVQANFSADPSAP-- 502
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 421 SPKLGWDEAMADPPNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHI-- 466
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 503 -GNLTFDSPITIDALFS---LGTSGIIFANTGPINTGNTLVTSPTNLDLFONKIQLNA 557
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 467 -----APWKGLETAYN---DLISPPS-----GAPLISALGDLG----- 498
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 558 NTFPTAPGYDIYFRKSVNGGFDLLGNANFYFDGAGITTPKLSFGVTATEIYVGNDIVT 617
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 499 -----GFTHADVT-AVPSITFTSPVFTTG-----TSVDYAEPLNP 531
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 618 QGNQIFDGVYGLQPVNLTSAGSVITFNILLNGSLQVQTAQNVISQPSSSLSAVEIAS 677
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 532 SIIVRAGSFDPSPQNDRHVAF---STDGKNMFQSGEPGGVTTG-----GTVA 577
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 678 DVLLNAG-----QNVSFGNINTRGGNVDIQAL--GNISTGSIIVTSFPGNAGNVI 725
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 578 ASADG-----SRFVAPGDPGPVYVAVFGNSWA-----ASQGV 612
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 726 LNAGGTLTTGYIETSGTNGGDDVTTSSGNTSTAYIDTRGFGDGLDLSLGGAVSIESKGD 785
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 613 PANAQIFSRVRNPKTFVALSNGTIFYRSTDGCVTFQPPVAAGLPSSGAVGVNMFHVPCKE-- 670
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 786 ITTAFIDTGAYSIESFNEGTGGNVLFTADGSITNTYIFTAGKNGG--DIFFQAGESIEII 843
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
QY 671 -----GDLWLAA-----SSGLYHSTNGGS-----SWSAITGV 697
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 844 DYLNTYGSQTSQSDGVYVEAPLDISIGSYIYGGGEPGNVFLQAGGDITTSYIDTSAANGG 903
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 698 SSAVNVG-----FGKSAPGSSYPAVFVVG 722
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 904 DIFQSGGDEVGYLFTKGYEGRGDGVYVETGRYFRAIDGFLICEGPFVYTAGLTVG- 962
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 723 IGGVTGAYRSDDCGTTWVLINDQHOYGNNGQAITGTHANLRRVYIGTNGR--GIVYGDI 780
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 963 -----GSVYI-----QFGGSEPFIIIGNPI-----TNTIGAISGDD 994
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 781 GGAPSGSP-----SPSVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 825
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 995 NTPVIGTPIPIFDITLDNIITITEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1054
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 826 PSFSPSPSPSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1055 PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1097
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 13
TI0361
hypothetical protein 92 - Orgyia pseudotsugata nuclear polyhedrosis virus
C;Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNVP
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10361
R;Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.
Virology 229, 381-399, 1997
A;Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedro
A;Reference number: Z17011; MUID:97271300
A;Accession: T10361
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-279 <HR>
A;Cross-references: EMBL:U75930; NID:g2934903; PIDN:AAC59091.1; PID:g1911338
C;Superfamily: proline-rich protein 3

Query Match          5.7%; Score 293; DB 2; Length 279;
Best Local Similarity 55.8%; Pred. No. 1.1e-06;
Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;

QY 783 APGSPSPSPSVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 842
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 78 SPRTPTSLSPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 137
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 843 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 138 SPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 163
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 14
S50754
hypothetical protein WP6 - Chlamydomonas eugametos
C;Species: Chlamydomonas eugametos
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C;Accession: S50754
R;Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Kils, F.M.; Goodenough, U.W.; Hari
Plant Mol. Biol. 26, 947-960, 1994
A;Title: Domain conservation in several volvoclean cell wall proteins.
A;Reference number: S50754; MUID:95093034
A;Accession: S50754
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-351 <WOE>
A;Cross-references: EMBL:L29028; NID:g530877; PIDN:AAB53954.1; PID:g530878

Query Match          5.7%; Score 291; DB 2; Length 351;
Best Local Similarity 64.4%; Pred. No. 1.8e-06;
Matches 58; Conservative 10; Mismatches 18; Indels 4; Gaps 2;
```


QY 783 APSCSPSPVSPASPSLSPSPSPSPSPSP--SPSSSPSPSPSPSPSPSPSPSPSPSPSP 840
Db 175 SP 234
QY 841 SASP--SPSSSP 868
Db 235 VASPOQSP 264

RESULT 15
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
A:Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GET>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: x
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 5.7%; Score 290.5; DB 2; Length 2232;
Best Local Similarity 21.8%; Pred. No. 1.6e-05;
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;

QY 18 SLLAATASFAVAALGVLPTA---ITASPAHAATTQPYTHSNVAI--GGGFFVDGIVFNE 72
Db 235 SLGTSSSLPSSISTALPIASSSSSSPAASSTPVLSSSTIQSSSGTTFPSSVASSP 294

QY 73 GAPGILYVRTDIGMYRWDAANGRWIPLLDWGVNWNMGVGVSIAD-----PINTNK 126
Db 295 STVG-----STSG-----AASSSYATVSTIAGSTGSTITPVPSS 330

QY 127 VWAAVGYMTNSWPDNDGAILR--SSOGATWQITP-LPFKLGGNMP---GRGMGERLAVD 180
Db 331 --STIGSSTPSASSSSSGTMTSTIGSTGTVTVVPGSSSTFASSTPIASSSSPGSTVTA 388

QY 181 PNNDNILYFGAPSGKGL---WRSTDGATWSOMTNPDPVCTYIANPTDTTGYOSDIOGVV 237
Db 389 PGSSSTYGSSTPSASSSSSGTMTNSGSGTSTVTVAPVSSSTFGSSTPIASSSSSGSTVT 448

QY 238 WVAFDK-----SSSLGQAS-----KTFVGVADPNPNPVFVWSRDGGATWQA 278
Db 449 VVSGSSSTYGSSTPSASSSSAGTASTIGSTGSTATIVPGSS-----SSVGSSTQSA 500

QY 279 VPGAPTCFIPKGVDFPDVNVLIATNTGCPYDGGSDGVKFSVGTGWTIRISPVPESTD 338
Db 501 SPSSPG-----TMTSTVSGP-TGSTVTVVPGSSSTSPA-PSSSPNPSSS 540

QY 339 TANDYFGYGLTIDRQHPNTIMVATOISWMPDTHFRSTDGATWT---RIWDVTSYPNR 395
Db 541 PAS-----TGSTITIGSSSIIIVST-----VSGSTVSGSTGTSQSTLASSTATPGS 586

QY 396 SLRYVLDISAE-----PWLTFGVQPNP---PVPSFKLGWMDMAIDPFNS 438
Db 587 SSTVPSSSPQSPSPAPNTGTTPTSQSSQSPSPSMNPSSSTPTGSSQSTITPEGSTA 646

QY 439 DRMLYGTGATLYATNDLTWKDSGGQHIAPMWKLEETAVNDLISPPSGAPLISALGDL- 497
Db 647 SSTGTGSGTFSVATEVTSQST-----VPSGSSSLGTQSTNNSPSPSSLSPSTSGMSTLT 700

QY 498 -----GGTFHADVTAVPS-----TIFTSPVFTTGTSDVYAEINLPSIIVRAGS 539

Search completed: July 2, 2002, 09:13:57
Job time: 125 sec

Db 701 SEPSPSSTOSSGAQSTLTTPSPNPSPSTSSLESSTSGAITSAGTTMTSPSQSSVGS 760
QY 540 FDPSSQPNDRHVAFTDGGKNWFQGS--EPG-GVTTGGTVAASADGSRFVWAPG----- 590
Db 761 SQGSTSP-----AASITSGEMTSQGSTQTTPGSSVSTSAAILTSTQOSVSTNSPGSTVTRP 815
QY 591 -----DPCQPVVY-----AVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSNGTFYRS 639
Db 816 STVSGSTSSGSTVTVGSTEASTSGSSVASSPAPSTSQ-----NPNPSTSSGSSMITQS 869
QY 640 TDGGVTQPVAAGLPSSGAVGVMFHAPGKGDWLAAASSGLYHSTNGGSSWSA-----IT 695
Db 870 PYPQSQTSPVESSTTFS-----PGSPGTLTSTSPSPSQSTIGTQGSTSPGIS 919
QY 696 GVSSAVNVGFGKSPAGSSYPAVFVVGITGGVTGAYRSDDCGTTWVLINDDQHOYGNWQA 755
Db 920 TTSEEMTSQGSTQTPGSTGTVTPQSTVSDST----- 951
QY 756 ITGDHANLRVYIGTNGRGIVYGDIGAGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 810
Db 952 -----SSGSTVTVGSTE-----SSSPIPSTQNTNPSTSSGSSMSTQTPQSSQ 995
QY 811 SPSP--SPSSSPSSSPSP-----SPSPSPSP-----SRSP-----SPSASPSPSS 848
Db 996 STSPVESSTGATSSSGSPGTTLTSPSPSPSPSTTGSSQSGSTSPVVSTISQGSTETPGS 1055
QY 849 SPSPSSSP-----SSSPSTPSSSPVSGGVKVVQYKNN 881
Db 1056 TGTVTNKPSTVSGSSGSTATMGSTEASTSGSSTSPNPQSSTSPSTSGA-----T 1108
QY 882 SAPGDNQIKPGLQVNTGSSVDLSTVTVRYWTRDGGSTLVYNC--DMAAICCGNIRA 939
Db 1109 SSPGSS-----GTTLTSTSPSPSQSSSTIG-----SSOGSTSPVVSTTSGDMTSOGSTQIPG 1159
QY 940 SFGS--VNPATPTADT 953
Db 1160 STGSTVTPQSTGSGST 1175

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 28.05 Seconds
(without alignments)
1321.020 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	397.5	7.7	1742	1 GUNA_CALSA	P22534 caldocellum
2	340.5	6.6	1331	1 MANE_CALSA	P22533 caldocellum
3	293	5.7	279	1 Y091_NPVP	O10341 orgyia pseu
4	290	5.6	1039	1 GUNB_CALSA	P10474 c endodluca
5	283.5	5.5	474	1 VTP3_TTV1V	P19275 thermoprote
6	272	5.3	555	1 GPL_CHURE	Q9f9q6 chlamydomon
7	261.5	5.1	1853	1 CIPA_CLOTM	Q06851 clostridium
8	250	4.9	268	1 NO20_MEDTR	P93329 medicago tr
9	245	4.8	772	1 CIPB_CLOTM	Q01866 clostridium
10	245	4.8	879	1 GUN1_CLOTM	Q02934 clostridium
11	244	4.8	1664	1 SLPI_CLOTM	Q06852 clostridium
12	235.5	4.6	2124	1 PGCA_RAT	P07897 rattus norv
13	222.5	4.3	360	1 VTPX_TTV1	P19274 thermoprote
14	221	4.3	2132	1 PGCA_MOUSE	O61282 mus musculu
15	218.5	4.3	3164	1 TEGU_HSV11	P21920 herpes simp
16	216	4.2	449	1 APG_BRANA	P40603 brassica na
17	214	4.2	532	1 SPG7_DICDI	P22698 dictyosteli
18	210.5	4.1	562	1 GUN1_ACICE	P54383 acidothermu
19	209	4.1	535	1 SPKC_SYNY3	P16112 homo sapien
20	209	4.1	2415	1 PGCA_HUMAN	P21997 volvox cart
21	208.5	4.1	485	1 SSGP_VOLCA	Q06885 dictyosteli
22	207.5	4.0	544	1 GPI0_DICDI	P40602 arabidopsis
23	205.5	4.0	534	1 APG_ARATH	P29719 paenibacill
24	205.5	4.0	700	1 GUNA_PAELE	P10475 bacillus su
25	204.5	4.0	499	1 GUN2_BACSU	P24928 homo sapien
26	204.5	4.0	1970	1 RPBI_HUMAN	P08775 mus musculu
27	204.5	4.0	1970	1 RPBI_MOUSE	Q59394 erwinia car
28	204	4.0	444	1 GUNN_ERWCA	P07983 bacillus su
29	203.5	4.0	499	1 GUN1_BACSU	P26414 microbisp
30	201	3.9	456	1 GUNA_MICBI	P23549 bacillus su
31	199.5	3.9	499	1 GUN3_BACSU	P08640 saccharomyc
32	199.5	3.9	1367	1 AMVH_YEAST	Q9y4h2 homo sapien
33	199	3.9	1324	1 IRS2_HUMAN	

ALIGNMENTS

RESULT 1

ID	GUNA_CALSA	STANDARD;	PRT;	1742 AA.
AC	P22534;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)			
DE	(Cellulase A).			
GN	CELA.			
OS	Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Thermoanaerobacter group; Caldicellulosiruptor.			
OX	NCBI_TaxID=44001;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=95336703; PubMed=7612247;			
RA	Te'O V.S., Saul D.J., Bergquist P.L.;			
RT	"cela", another gene coding for a multidomain cellulase from the			
RT	extreme thermophile Caldocellum saccharolyticum.;			
RL	Appl. Microbiol. Biotechnol. 43:291-296(1995).			
RN	[2]			
RP	SEQUENCE OF 1516-1742 FROM N.A.			
RP	MEDLINE=91247819; PubMed=2039230;			
RA	Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;			
RT	"Cloning, sequence analysis, and expression in Escherichia coli of a			
RT	gene coding for a beta-mannanase from the extremely thermophilic			
RT	bacterium 'Caldocellum saccharolyticum'.			
RL	Appl. Environ. Microbiol. 57:694-700(1991).			
CC	-I- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN			
CC	ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL			
CC	DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE			
CC	CELLULOSE.			
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-I- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL			
CC	SITE FOR PROTEOLYSIS.			
CC	-I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	E (FAMILY 9 OF GLYCOSYL HYDROLASES).			
CC	-I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY			
CC	L (FAMILY 48 OF GLYCOSYL HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L32742; AAA91086.1; -			
DR	EMBL; M36063; AAA72860.1; -			
DR	EMBL; L01257; -; NOT_ANNOTATED_CDS.			
DR	PIR; A43745; A43745.			
DR	HSSP; P26221; 1TF4.			
DR	InterPro; IPR001956; CBD_3.			

Q47096 erwinia car
P11414 cricetus
P16356 caenorhabdi
Q28343 canis famil
O13368 candida alb
P12021 sus scrofa
P46590 candida alb
P35084 dictyosteli
P35088 bos taurus
P02893 plasmodium
P18616 arabidopsis
P19597 plasmodium

DR InterPro: IPR000556; Glyco_hydro_48.
DR InterPro: IPR001701; Glyco_hydro_9.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR00844; GLHDLRLASE48.
DR ProDom: PD001947; CBD_3; 2.
DR ProDom: PD011903; Glyco_hydro_48; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1742 ENDOGLUCANASE A.
FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1742 CATALYTIC 2.
FT ACT_SITE 396 396 BY SIMILARITY.
FT ACT_SITE 434 434 BY SIMILARITY.
FT ACT_SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T -> A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;

Query Match 7.7%; Score 397.5; DB 1; Length 1742;
Best Local Similarity 21.7%; Pred. No. 4.9e-10;
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

QY 262 PNNPVFWSRGGATWQA-----VPGAPTGFPHKGVDPVNHVLY-IATSNLTGPGYDGSSG 316
DB 48 PN-----NWRGDSALKDQDNGLDLTGWFEDAGDHVFNLPMSYTGTLMSWAAY 100

QY 317 DVNKFVSTGTRISPVSTDTANDFYGSGLTIDRQHPNTIMVATQI-----SWW- 368
DB 101 EYKDAFVKSGLEHI--LNOIEWNDYF-----VKCHPSKVVYVYQVGGDKDHAWWG 151

QY 369 -----PDTIIFSTGGATWTRIDWTSYPNSRLRVLDISAEPWLTFGVQNPVPV 419
DB 152 PAEVMQMERPSFKVTQSSPGSA-----VVAETAASLAASASIVILKDRN 193

QY 420 PPSKLGMDMAIDAIDFNSDRMLYGTATLYATNDLTKWDSGQIHIAPMKVGLLEETAVN 479
DB 194 PTRAAATYLAQHA-----KDLYEFAEVTKSDSG-----YTAAN 224

QY 480 DLISPPSG-----APLISALDGLGFTHADVTAVPSTFTSPVFTTGTSDVY---- 526
DB 225 GYVNSWGSFYDELSWAAVWLYLATNDSTYLTAE-----SYVQNPWKISGNSNLDYKWAH 279

QY 527 -----AELNPSIIIRAGSEDPSSQPNDRHVAFSTDG-----GKNW----- 561
DB 280 CWDVHNGAAL-----LKAITDKDTYKQIIESHLDYWTGTGNGRIKYYTKGLAWLDQWS 336

QY 562 -----FOGSEPGGVYTTGGTVAASADG--SRFVWAPGDPQPVYAVGFG----- 603
DB 337 LRYATTATLAFYVDSWSCPTCKKTYRKFGESQIDYALGSTGRS--FVWGFETNPKR 394

QY 604 -----NSWAASQGVPA-----NAQIRSD----- 621
DB 395 PHRTAHSSWADSSQPSVYHRHTLYGALVGGPGSDSDSYTDDISYVNVNEVACDYNAGFVG 454

QY 622 -----RVNP-----KTFYALNSGTFYRSTDDGGYTFQPVAGLPSSGAVGVMFHAVPG 668
DB 455 ALAKMYLLVGNPNIPDKAETETNDEFF-----VEAGINASGTFNTEIKAI-- 501

QY 669 KEGDLMAASSGLYHSTNGSSWSA-----ITGVSSAVNVGFGKSAPGSSYPAVE 718
DB 502 -----VNNQSGWPATNKLKFRYFVLDISELLKAGYS,-----PNQL 537

QY 719 VVGTYI-----GGVTGAVRSDCGTGTWVLIND-----DQHQYGNWGQAITGDHANLR 765
DB 719 VVGTYI-----GGVTGAVRSDCGTGTWVLIND-----DQHQYGNWGQAITGDHANLR 765

538 TLTNTYNOGAKVSGPYWMDSSRNYYILVDFGTGLIYPGGQDKYKKEVQFRIAPQNVQ- 596
766 VYIGTNGRGIVYGDIGGAPSGS-----PSPSVSPSASPSLSPPSPS 807
597 ---WDNSNDYSFQDIKGVSSGVVKYIPLYDEDIKWGEPEGTS-----GVSPPTAS 648
808 SSPSS 867
649 VTPPTPTPTA-----TPT 700
868 P-VSGGVKVOYKNDAPGDNQIKPGLQVNTGSSSSVDLSTVTVRVWFTRDGSSSLVYN 926
701 PATSGQIKVLYANKETNTTIRPWLKVNSGSSSIDLSRVTIRYWTYVDGERAQSAS 760
927 CDWAAITCGGNIRASFGSVNPAATADTYLQ 956
761 -DWAQIGASVNTPKFKVLSVSSVSGADYILE 789

RESULT 2
MANB_CALSA STANDARD; PRT; 1331 AA.
ID MANB_CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=93119139; PubMed=1476429;
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
multidomain enzyme."
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luechi E., Jasmal N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
gene coding for a beta-mannanase from the extremely thermophilic
bacterium 'Caldocellum saccharolyticum'";
RT Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
linkages in mannans, galactomannans, glucomannans, and
galactoglucomannans.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -!- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CELSIUS.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
J (FAMILY 44 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; L01257; AAA71887.1; -

[illegible]

	Query Match	4.8%	Score 245;	DB 1;	Length 772;
	Best Local Similarity	24.4%;	Pred. No. 0.0006;		
	Matches	95;	Conservative	68;	Mismatches 116; Indels 110; Gaps 14;
QY	634	GTFYRSTDDGGVTPQPVAAIGLPSSGAGVMHFAPVG-KEGDWLAAASGLYHSTNGGSWS	692		
		: : : : : : : :			
Db	28	GTYAITODG--VFATIVATVKSAAAPITLLEVGAFRDNDL-----VEIST-----	71		
		: : : : : : :			
QY	693	AITCVSSAVNVGCKSAPGSSYPAVFVCTIGGVGTGAYRSDDCGTTWVLINDDQHQTGNW	752		
		: : : : : : : : :			
Db	72	--TFVAGGVNLGSSVPQTQPNSVDGVIIEGKVTGS-----VGTT---VEIPVYFRGVP	121		
		: : : : : : :			

Qy	753	GOAITGDHANLRVY-----IGNNGRGIVYDIGGAPSGSPSPSVS-----	793
Dd	122	SKGI-----ACDFFERYDPNVLEIIGIDPRSIIV-----DPNPTKSFDTAIYADRK	168
Qy	794	-----PSASPSLSRPPSSPSSPSPPSPSSPS-----	821
Dd	169	IIVFLFCGRNRNSVSITKGVFAKIRATVKSSAPAYITFDEVGFGADFNDLVEOKVSFID	228
Qy	822	-----SSPSPSPSPSRPSRSRPSRPSRPSRPSRPSRPSRPSPTSPSSPVSGGVKVQ	876
Dd	229	GCVNVGNATPKGATPNTATPTKSA---TATPGHVPINTPTANTPYVSGNLKVE	284
Qy	877	YKNDSAPGDNQIKPGLQVVNWGTSSVDLSTVTYRVYFTRDGGSTLIYNCDMAAI----	932
Dd	285	FYSNPSDDTNINSINPFKFVNWTGSADLSKLTRYTYTTVDQGDKDTFW-CDHAALIGSN	343
Qy	933	-----GCCNTRASEGSVNTPATTADTYLO	956
Dd	344	GSYNGITSNVKGTFVKMSSSTNNADTYLE	372
RESULT 10			
GUN1_CLOTM STANDARD; PRT; 879 AA.			
ID	GUN1_CLOTM	SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.	
AC	Q02934;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)		
DE	(Cellulase I).		
GN	Celi.		
OS	Clostridium thermocellum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium		
OX	NCBI_TaxID=1515;		
[I]			
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.		
RC	STRAIN=NCIB 10682;		
RX	MEDLINE=G3171873; PubMed=8436949;		
RA	Hazlewood G.P.; Davidson K.; Laurie J.I.; Huskisson N.S.;		
RA	Gilbert H.J.;		
RT	"Gene sequence and properties of Celi, a family E endoglucanase from Clostridium thermocellum."		
RL	J. Gen. Microbiol. 139:307-316(1993).		
CC	-!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-		
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-		
CC	GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLOCAN.		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.		
CC	-!- PATHWAY: CELLULOSE DEGRADATION.		
CC	-!- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL HYDROLASES).		
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licensee@isb-sib.ch).			
CC	EMBL; L04735; AAA20892.1;		
DR	PTR; A47704; A47704.		
DR	HSP; P26221; ITF4.		
DR	InterPro; IPR001956; CBD_3.		
DR	InterPro; IPR001701; Glyco_hydro_9.		
DR	Pfam; PF00942; CBD_3; 2.		
DR	Pfam; PF00759; Glyco_hydro_9; 1.		
DR	ProDom; PD001947; CBD_3; 1.		
DR	PROSITE; PS00582; GLYCOSYL_HYDROL_F9_1; 1.		
DR	PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.		
KW	cellulose degradation; Hydrolase; Glycosidase; Signal.		
FT	SIGNAL		

FT	SIGNAL	1	55
KW	Cellulose degradation;		

FT	DOMAIN	486	580	G2-B.
FT	DOMAIN	587	682	G2-B.
FT	DOMAIN	685	798	KS.
FT	DOMAIN	801	1226	CS-1.
FT	DOMAIN	1227	1909	CS-2.
FT	DOMAIN	1910	2124	G3.
FT	DISULFID	51	133	BY SIMILARITY.
FT	DISULFID	175	246	BY SIMILARITY.
FT	DISULFID	199	220	BY SIMILARITY.
FT	DISULFID	273	348	BY SIMILARITY.
FT	DISULFID	297	318	BY SIMILARITY.
FT	DISULFID	509	580	BY SIMILARITY.
FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1914	1925	BY SIMILARITY.
FT	DISULFID	1942	2034	BY SIMILARITY.
FT	DISULFID	2010	2026	BY SIMILARITY.
FT	DISULFID	2041	2084	BY SIMILARITY.
FT	DISULFID	2070	2097	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1842	1842	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;		

Query Match 4.6%; Score 235.5; DB 1; Length 2124;
Best Local Similarity 21.3%; Pred. No. 0.0044;
Matches 224; Conservative 115; Mismatches 359; Indels 353; Gaps 48;

QY	112	NGVSTAAADPINTNKVAAWMTNSWDPNDAIGAILSSDOGATWQITPLPFKLGGMNPR	171
Db	979	SGIEDISVLTGPSTSPETASGVGDLSGLPSGGESLETASGVF-DVTQLPTERRGLETSA	1037
QY	172	GMGERLAVDNNNIL-----YFGAPSGKGLNRSTDGATWSQMTNFPD-----	215
Db	1038	SGIEDITVLTGRENLETASGVDEDVSLGPSGKELETASGI--EDISVFPEAGLET	1095
QY	216	--VGTIYAN-PTDTTGYSDIQVWVAFDKSSSLQSASKTIFVGVADPNPNVFWNRDG	272
Db	1096	SASGGYVSGIPSGEDGTETSTSGVEGVS---GLPSGEGLETASGVDELGLPTRDSLET	1152
QY	273	GATWQAVPGAPTC-----FIPKHGVDPVNHVLY---IATSNVG-----GPYD-	312
Db	1153	SASGVDTGTYPGREDETETSP--GVGDDLSGLPSGOEGLETASGAEDLGLGPSKEDL	1210
QY	313	-GSSGDVWKF-----SVTSGTWTTRISPVPTDNDYFG-----YSGL---	349
Db	1211	VGASGALDGLKPLSGTGLSCQTPEASGLPS-GFSEYSDVDIGSGPSSGLPFGSLPSG	1269
QY	350	--TIDRHPNTIMVAT-----QISWMPDTIIFRSTD--GGATWTRI	386
Db	1270	FPTVSLVDSTLVEVITATTASELEGRTISVSGSGEESGPPSLSELDSSADISGLPSGTEL	1329
QY	387	WDMTSYPNRSLRYLVDISAPFWITFGVQPNP-----PVPSPKLGMW	427
Db	1330	SGOTSGS-----LDVSGTSGTFDVSQGPFGSGGTGEGTSGIPEVSGQAVRSPDTEI	1382
QY	428	DEAMADPFNSD-----RMLYGTGATLYATNDLTKWDSGGQIHAPMVKGLEETAVNDL	481
Db	1383	SELSGLSSGQPDVSGEGSLIFGSGSGSGLTS-----VSG-ETSGISDL	1425
QY	482	ISPPSGAPLISAL-----GDLGFTHADVTAVPSTIFTSPVFTTGTSDVY	526
Db	1426	SGQPSGFPVLVSGTTPGTPTDLASGMSGSDSGITFVD-----TSLIEVTPTFRE	1476
QY	527	AELNPSIIVRAGSFDPSOPNDRHVAFTDGGKNWQSGPEPGVTTGGTVAASADCSRFX	586
Db	1477	EE-----GLGSVELSGULPSGGETDLSGTSGMYD-VSGOSSGAIDSSGLISPTPEFS---	1525

QY	587	WAPDCPQPVVYAVGFGNSMAASQGVYAN-AQIRSDRVNPKTFYALSNGTFYRSTDC---	642
Db	1526	-----GLPSGVAEVSGEVSGVETGSSLSGAF-----DGSGL	1557
QY	643	-----GVTFQVPA--AG-----LPSSGAGVGMFHAVPKGKGLD-----	673
Db	1558	VSGFPTVSLVDRTLVESITLAPTAQEAEGEPSSILEFSGA---HSGTPDISGDLGSLD	1613
QY	674	-----WLAAS-----SGLYHSTN--GGSSNSAITGVSSAVNVGCKAPGSSYP	715
Db	1614	QSTWQPCWTEASTEPPSSPYFSGDFSSTTASGESITAPTG-----SGETSGLP	1662
QY	716	AVFVGT--IGVGTGAYRSDCGTTWLVINDQHOYGNMQAITGDHANLRRVYIGTNGR	773
Db	1663	EVTLITSELVEGTEPTVSOELG-----HGPSMT-----YTPRLF-EASGE	1702
QY	774	GIVYDIGG-----APSGSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSP	820
Db	1703	ASAGDLGGPVTTFPGSGVEASVPEGSSDPSPAYPEAGVGVSAAPSAEQSLSEFPLHGIT	1762
QY	821	SSS-----PSP	869
Db	1763	SASRETDLMTTPTCTEVSSNPWTFQETRGSAAPESVSGESTSDIDACTSGVPFATPM	1822
QY	870	SGGVKQYKNNDSAPGDNQIKPGLQVNVNTGSSSVDLSTVT--VRYWFTRDGSGSTLVYNC	927
Db	1823	TSQDRTEISGWSND-----HSEVNVVTVTTPESRWAQSTOHPTELT---	1865
QY	928	DWAAICGNIRASFGSVNPA-----TPTADT	953
Db	1866	-----QEIGSPNPSPSYGEEQTQATET	1885

RESULT 13
ID VTPX_TTV1 STANDARD; PRT; 360 AA.
AC P19274;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain KRAL) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC Lipothrixvirus.
OX NCBI_TaxID=10480;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174928; PubMed=2308830;
RA Neumann H., Zillig W.;
RT "The TTV1-encoded viral protein TPX: primary structure of the gene and the protein.";
RL Nucleic Acids Res. 18:195-195(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14855; CAA33002.1; -.
DR PIR; S12850; S12850.
KW Repeat.
FT DOMAIN 278 353 3 THR-PRO REPEATS REGIONS AND TWO NEAR IDENTICAL REPEATS.
FT REPEAT 270 291 THR-PRO(N).
FT REPEAT 292 301
FT REPEAT 302 322 THR-PRO(N).
FT REPEAT 323 332
FT REPEAT 333 353 THR-PRO(N).
SQ SEQUENCE 360 AA; 38214 MW; EF54FBCB753ED1CA CRC64;

FT DOMAIN 1917 2132 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 1922 1933 BY SIMILARITY.
FT DISULFID 1950 2042 BY SIMILARITY.
FT DISULFID 2018 2034 BY SIMILARITY.
FT DISULFID 2049 2092 BY SIMILARITY.
FT DISULFID 2078 2105 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 1171 1173 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2132 AA: 222008 MW; 0B2BCDFC6C8DA163 CRC64;

Query Match 4.3%; Score 221; DB 1; Length 2132;
Best Local Similarity 20.9%; Pred. No. 0.018;
Matches 199; Conservative 102; Mismatches 311; Indels 338; Gaps 46;

QY 110 GYNGVWSIAADPINTNKVAAVGMYNWDNDGAILRSSDQGA-----TWQITPL 160
DB 1105 GYVSGIPSGDGTET-----SASGVEDVSLPGSGEGLETSASGVEDLGPSTRDSLETAS 1160
QY 161 PFLLGGMPCRGWGERLAVDPNDNDILYFCAPSKGLWRSTDSGATWSQNTFPD----- 215
DB 1161 GVDVTGPPSGRGPETSVSGVGD-----FSLGPSKGELETSSASGA---EDLSGLPSGKEDL 1215
QY 216 -----VGYIANPT-DTTGYQSDIQGVVWAFDRSSSLGQASKTIFGVGA 260
DB 1216 VGSASGALDFGLKPLPGTLGSGQTPPEVNGPFGSG-----EYSGADIGSGPSS---GLP 1266
QY 261 DPNNPFWSRDGGATWQAVPGATGTFPHKGVFDPVNVHLYIATSNITGGPYDG----- 313
DB 1267 D-----FSLGPSGF-PTVSLVD---STLVEVITATTSLEGRGTIGIS 1306
QY 314 SSGDVWKFVTSCTWTRISVPSTDTANDYFGYSGLTIDRQHPNTINWATQISWMPDIT 373
DB 1307 GSEV-----SGL-----PLGELDSSAD---ISGL-----PSGTELSQASGSPDS-- 1344
QY 374 FRSTDGGATWRIWDWTSYPNRSLRYVLDISAEFWLTFFGVQPNPPVPSPKLGMDEAMAI 433
DB 1345 SGETSG-----FFDVSGQPFQSSGVY-----EETSGI 1371
QY 434 DPNSDRMLYGTATLYATNDLTKWD---SCGQTHIAPMWKGL-----EET 476
DB 1372 PEISGQPS--GTDPTT-ATSGVTELNELSSGQPDVSGDGLFGSGQSSGITSVSGETS 1428
QY 477 AVNDLISPPSGAPL-----ISALDGLGFTHADVTAVPSTIFT----- 514
DB 1429 GISDLSCQPSGFPFVSGTATRTFDLASGTISGSGSGITFVDTSFVEVTPPTFREEEGL 1488
QY 515 -----SPVFTTGTSDVYAEALNPISIIVRAGSFDPSQPNDRHVAFTDGGKNWF 562
DB 1489 GSVELSGFPSGTELSGTCTVDVSEQSGAIDSSGLTSPTPE-----F 1532
QY 563 QGSEPG-----GVTTGCTVAASA-DGSRFWAPGDP-----GQPVVY 598
DB 1533 SGLPSGVAEVSGBFSGVETGSSLSPLSGAFDGSGLV--SGFTVSLVDRTLVESITQAPTAQ 1590
QY 599 AVGFGNISW-----AASQGVPANAIQRSDRVNPKTFYVSLNSGTFYRSDGGVTFQPVAAAGL 653
DB 1591 EAGEGPGSILEFSGAHSQGP---DISGELSGDLSTLQSGQMETSTE-----T 1636

QY 654 PSSGAVGVMEHAPVKGKEDLWLAASSGLYHSTNGGSSWSALTGVSSAVNVGFGKSAFGSS 713
DB 1637 PSS-----PYFSGDF-----SSTTVSGESIAATTTGSGES---SG 1668
QY 714 YPAVFVVGT--IGVGTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLRVYIG-- 769
DB 1669 LPEVTLNLTSELVEGVTEPTVYSELG-----HGFSMT-----YISRL 1704
QY 770 --TNGRCIVYDGGAPSGSPFSVS---PSASPSLSPPSPSSSPSSSPSSSPSSSP 824
DB 1705 SEASGDASASGDLGAVTNFPGSGVEASVPEASDLSAYPEAGVGVGAAPASAKLSEFP 1764
QY 825 -----SPSPSPSPSP-----SRSPSPSPSPSPSSSPSSSPSSSPSSSP 863
DB 1765 DLHGITSAPFHTDLEMTTPSTEVSNSNPWFQEGTREGSAAPEVSGESTTSDITGTSGV 1824
QY 864 PSSSPVGGVKKVQYKNNDSAPGDNQIKPQLQVNVNTGSSVLDLSTVTVRYW 913
DB 1825 PSATPMASGDRTEISGEWS---DHTSE-----VNVAIS---STITESEW 1862

RESULT 15
TEGU_HSV11 STANDARD; PRT; 3164 AA.
AC PI0220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Viroin protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1";
RL J. Gen. Virol. 69:1531-1574(1988).
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EBV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14112; CAA32311.1; -
DR PIR; I30085; WMBEH6.
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA; 335857 MW; CC5D31FF4F9FE3F4 CRC64;

Query Match 4.3%; Score 218.5; DB 1; Length 3164;
Best Local Similarity 20.3%; Pred. No. 0.035;
Matches 236; Conservative 100; Mismatches 402; Indels 427; Gaps 58;

QY 24 ASFAVAAALGVLIATASPAHAATTPQYTWSNVATGGGFGVDG-----IVFN 71
DB 2074 ASDAAAEALYVALVATTLTREFGC-----RWAQLEMGWGDAAAFPPGLPQSGSTRHRVSFN 2128
QY 72 EGAPGILYVTRTDIGGVRWDAAANGRWIPLLDWVGWNNWGY---NGVYSIAADPINTNKV 128
DB 2129 ENDVLVALVASSPEHIYTF-----W--RLDLVRQHEYMHLLTPRAFQNAADSW----- 2174

Search completed: July 2, 2002, 09:17:53
Job time: 361 sec

```
QY 129 AAVCMYTNWDNDGAILRSSDQGATWQITPLPKLGGNMPGRGM--GERLA----- 178
Db 2175 ---LFFVORLTPHDA-----RIRVLPFAFSAGPPTRLMGFTRLADWRRGKLS 2219
QY 179 -VDPNND--NILYFAPSGKGLWR-----STDGATWSOMTFPDV 216
Db 2220 ETDPLAWRSVPELGTGGAALGKLSPAQALAAVSVLGRMCLPSTALVALWTCM--FPD- 2276
QY 217 GTYIANPTDTGYSODIQGVVWVAFDKSSSLGOASKTIFVGVADPNPNPVFWSRDGG--A 274
Db 2277 -----DYTEYDS-----FDALLTARLESQOTL-----SPSGGREA 2306
QY 275 TWOAVPGA---PTGFIPHKGVDFDPVNH----- 298
Db 2307 SPPAPPNALYRPTG--OHVAVPAATHRTPAARVTAMDVLAAVLLGAPVVVALRNTAF 2364
QY 299 -----VLYIAFSNTGGP-----YDSSGDVMKFSVTSGTW-----TRISPVSTDTA 340
Db 2365 SRESELELCLTLFDSRARGPDAALRDVAVSDI-----ETWAVRLLHADLNPIENACLA 2417
QY 341 NDYFGYSGLLTIDR--QHPNTIMVATQISWMPDTIIFRSTD--GGATWTRIMDWTSYPNRSL 397
Db 2418 AQLPRLSALIAERPLARGPPCLVLVDISMTPTVAVLWENPDPPGPDVRFVG--SEATEEL 2475
QY 398 RYVL--DISA-----EPWLTFCV-----QPNPPVPS 422
Db 2476 PFVAGGEDVLAASATDEDFLARAILGRPFDAISLLSGELFPGHPVYQRAPDDQSPSPVNP 2535
QY 423 KLGWMEAMA---IDPNSDRMLYGTGAT-----LYATNDLTKWDSGGQIHIA 467
Db 2536 TPGPVDLVGARGSLGPGSLAPTF--TQATGCEVPYPRMWAIHGLELELASDSDSGPAPLL 2594
QY 468 ---PMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGSV 524
Db 2595 APDPLSPTADQSVTSQCAPRPGPAVTAREARPGVPAESTRAP-----VGPRD 2644
QY 525 DYAEINPSIIVRAGSFPDSSOP-----NDRHVAFTDG-----CKNWFQ 564
Db 2645 DFRRL-PS-----POSSAPPDAPATPRPPASSRASAASSSGSRARRHRRARSILARA 2694
QY 565 SEPGGVITG-----GTVAASADGRFVWAPDGPQPVVYAVGVG-----NSWAASQGV 612
Db 2695 TQASATTQGWPPPALPDTVAPVTFAR--PPAPKPPPEPAPHALVSGVPLPLGPOAQAS 2753
QY 613 PANAIQRSDRVNKTFFVALSNGTYRSTDGGVTFPQVAAAGLPSSGAVGVMFHAYPGKEGD 672
Db 2754 PA---LPIDPVPP-----PVATGTV---LPGCENRRPPLTSGPAPTTPRV---PVGGPQR 2800
QY 673 LWLAASSGLYHSTNG-GSSW-----SALTGVSSAVNVGFGKSAPGSSYPAVFVVG---T 722
Db 2801 LTRPAVASLSRESLPSMDPADPTAPVLGRNPAEPTSSSPAGSPPPPAVQVAPVPPPT 2860
QY 723 IGGVTGAYRSDCCGTWVLINDQHXYGNWQAITGDHANLRRVYIGTNGRGIVYGDIGG 782
Db 2861 SGPPP-----TYLTLEG-----GV 2874
QY 783 APSG--SPSPSV-SPSASPSLS-----PSPSPSSSPSPSPSSSPSSSPSS 825
Db 2875 APGGVSRRTTRQPVATPTTSARPRGHLTVSRLSAFQPOQPOQPOQPOQPOQPOQPOQ 2934
QY 826 PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 885
Db 2935 PQQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 2987
QY 886 D-----NQIKPGLQV-----VNTGSSSDLSLSTVTVRYWTFTRDGGSTLVVNCWAA 931
Db 2988 EYPAVRFAQONRFSVPASASSNTPRTGSSLSGVSS-----W-----ASSLALHID--- 3033
QY 932 ICGNIRASFSGSVNPATPTADTYLQ 956
Db 3034 -----ATPPPVSLIQ 3043
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:11:52 ; Search time 91.42 Seconds
(without alignments)
1810.940 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTHRSRLVSL.....RASFGSVNPAIPTADTYLQX 957

Scoring table: BLOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	47.4	890	2	086727
2	2420.5	47.1	996	2	09AQH0
3	2037	39.7	839	16	097KK0
4	1784	34.7	806	3	09P4T8
5	1740	33.9	856	3	074170
6	1161	22.6	707	16	09WTE1
7	431.5	8.4	1779	2	052374
8	409	8.0	1770	2	09X3P5
9	402.5	7.8	921	2	09L8L8
10	398.5	7.8	1711	2	P96311
11	372.5	7.3	473	10	Q39620
12	350.5	6.8	997	2	Q92411
13	346.5	6.7	901	2	Q44562
14	334	6.5	1751	2	09AQG4
15	325.5	6.3	930	2	09RFX5
16	325	6.3	1000	2	Q24820

17	320.5	6.2	991	10	09AC44
18	318.5	6.2	875	2	09F2B0
19	308.5	6.0	611	16	P74375
20	301.5	5.9	1749	16	P73032
21	297	5.8	1915	2	Q9RPL0
22	291	5.7	351	10	Q39492
23	290.5	5.7	2232	5	P91365
24	286	5.6	261	2	09AQG7
25	279	5.4	170	2	09REF6
26	277	5.4	234	12	Q91GH4
27	276	5.4	1426	2	Q9X3P6
28	272	5.3	555	10	09FP06
29	265.5	5.2	955	10	094FP2
30	262	5.1	2468	16	Q912M3
31	258.5	5.0	554	6	Q95189
32	257	5.0	1325	5	Q9BKV7
33	254.5	5.0	913	2	Q59164
34	252.5	4.9	1498	2	Q9L448
35	252.5	4.9	2117	3	Q96U11
36	247.5	4.8	321	2	Q46392
37	246.5	4.8	818	6	Q9N1P0
38	245	4.8	887	2	Q9L3J8
39	243.5	4.7	552	10	Q96343
40	243	4.7	288	12	Q84565
41	236	4.6	1003	2	Q9AHP2
42	234.5	4.6	508	2	Q9FX3
43	233.5	4.5	3325	12	Q91BT9
44	233	4.5	873	5	Q9Y076
45	232.5	4.5	687	2	Q9KYF3

ALIGNMENTS

RESULT 1

086727 ID 086727 PRELIMINARY; PRT; 890 AA.
AC 086727;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED CELLULOSE.
GN SC5C7.30C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031515; CAA20642.1;
DR HSP; P07986; IEXG.
DR InterPro; IPR002860; BNR.
DR InterPro; IPR001919; CBD_2.
DR Pfam; PF02012; BNR; 10.
DR Pfam; PF00553; CBD_2; 1.
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Q94C44 chlamydomon
Q9f280 thiobacillu
P74375 synechocyst
P73032 synechocyst
Q9rpl0 acetivibrio
Q39492 chlamydomon
P91365 caenorhabdi
Q9aq7 caldicellul
Q9rfx6 caldicellul
Q9x3p6 caldicellul
Q9fpq6 chlamydomon
Q94f92 chlamydomon
Q912m3 pseudomonas
Q95189 bos taurus
Q9bkv7 leishmania
Q59164 actinomycet
Q91448 arthrobacte
Q96u11 neurospora
Q46392 clostridium
Q9nlp0 bos taurus
Q9l3j8 clostridium
Q96343 brassica na
Q84565 paramesium
Q9ahp2 arcobacte
Q9f2x3 streptomyce
Q9ibt9 turkey herp
Q9y076 leishmania
Q9kyf3 streptomyce


```
Db 472 PHLESALEDIAGFRHDELEKAPNWTYYQPNMGTTTDDFAELNPNFVRVNGVDKQNP 531
Qy 548 DRHAFSTDDGKNNFQ-SEPGGVTTGGTVAASADGSRFVWAPDGPQPVVYAVGFNSW 606
Db 532 TNRIGFSYDGGKSWFCQNETPEQGTSEGTVAADGSAVWAP--KCAKVCYSTDNGKW 589
Qy 607 AASGVFANQIRSDRNPNTFYALSNGTYRSTDDGVTF-QPVAAGLPSGGAGVGMFHA 665
Db 590 VECANVSEAAIVSDRNPNTFYALSNGTYRSTDDGVTF-QPVAAGLPSGGAGVGMFHA 645
Qy 666 VPGKEGLDLWLAASGLYHSTNGGSSAITGVSAVNVGFGKSPAGSSYPAFVVGTTG 725
Db 646 VPGLEGDLWLVNNGMWSHSDGYSFVKISGVEDAAISIGFGKPAEGTYPAIYYAKING 705
Qy 726 VTGAYRSDCGTWTWLINDDOHQYGNMGQAITGDHANLRVYIGTNGRIGVYDGGAPS 785
Db 706 VRGIFRSDCDKWTIRINDKHQGCANADITGPRVYGRVFVATNGLGKMGGEA--Y 762
Qy 786 GSPSPVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 845
Db 763 SNISPSATSTPSFTPTPTATITSTAPTSTPTPTPTPTPTPTPTPTPTPTPTPT 822
Qy 846 PSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 904
Db 823 MTPPTP-----TPTPTPTGTCGSLKLVLYKNNTSASAGSIRFWFKLVNGGSSVD 876
Qy 905 LSTVTVRYWTRDGGSTLVYNCDMAAIGCGNIRASFGSVNPARPTADTYLQ 956
Db 877 LSRVKIRYWTVDGKQSAV-CDMAIGASNTVTFEVLKSSGVGADYVLE 927

RESULT 3
Q97KKO PRELIMINARY; PRT; 839 AA.
AC Q97KKO;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PROBABLY SECRETED SIALIDASE, SEVERAL ASP-BOXES AND DOCKERIN
DE DOMAIN.
GN CAC0919.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hittl J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT *Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.*;
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007608; AAK78895.1; -
DR InterPro: IPR002860; BNR.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF02012; BNR; 9.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF01344; Kelch; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;
```

Query Match 39.7%; Score 2037; DB 16; Length 839;
Best Local Similarity 49.5%; Pred. No. 7.9e-95;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

```
Qy 13 SRRVLSLLA-----TASFAVAALGLVPLTAITASPAAHAAT--TOPYTWNSVAIGGGFVD 66
Db 2 NKRTVSVAGLSIIITGTF-----VTHSAAANKAAASVQSQYKWDNAKIGAGYVP 54
Qy 67 GIYFNECAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGWNNMGYNGVYSTAADPINTNK 126
Db 55 AVLFNKTEKOLLIVARTDDGGAYRWDKANNKIPLTD--GFSDMTLMGCESLATDIDTNR 112
Qy 127 VAAVGYNTSWPNDCGAILRSSDOGAQTWQITPLPFKLGNNMPCGRMGERLAVDPNNNI 186
Db 113 VYIAAGLYTNDWDENAYILLSSQDKNGTKRYQLPFKVGCMPCGRNMGERLQIDPNDKI 172
Qy 187 LYFGAPSGKGLNRSTDSGATWQNTNPPDVGTIANTDTTGTQSDTQGVVWVAFDKSS 246
Db 173 LYLGAARSGNLKMSYDGYGTSKVDNFPDPTGQVQDPQNE--YTADKVGVMWTFDFSTG 230
Qy 247 SLGOASKTIFVGVAD--PNNPVFMSRDGGAOWAVPGAPTGFPHKGVDFPNVHLYIATS 305
Db 231 TKSPPTMTVGAADKTGNNIYVNDGGKTVSAVKGPKGYLPHHGIL-ASDGLMILSYS 289
Qy 306 NTGSPYDSSGDVWKFVSTGWTTRISPVSTDTANDYFGYSLGLTIDRQHPNTIMVATQI 365
Db 290 NTCGPDYDSDGVWQYNTKGTGEWNTITPPAVGDTKS---GFGGISVDAQNNVNVVATLN 346
Qy 366 SWPPTTIFRSTDDGATWTRIDWTSYPNRSLRYVLDISAEPWLTFG-VQPNPPVPSPL 424
Db 347 RWPDEEIRYRSTDAKTKWPIWDMNGYPNRTLGYNLDYSAQPLDWDGKTGVTTPDPLVK 406
Qy 425 GWNDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDSCGQIHIAPMVKGLEETA VNDLISP 484
Db 407 GWMGDLIEDPFNSDRMFYGTGATLYGTDLTWWDKKNVDISVKANGIECAVNDVVVP 466
Qy 485 PSGAPLISALDGLGTFHADVTAVPSTIFTSTVDYAEINPNSIIVRAGSFDPSS 544
Db 467 TKGAQLLSAVGDDCGFYHDDITKVPKMTTTPNFSAITSDIYAEVSPFVVRVGNVDTSK 526
Qy 545 QPNDRHVAFSTDCGKNWFO--GSEPGVYTTGTVAASADGSRFVWAPDGPQPVVYAVGF 603
Db 527 NQDKCGISYDGGKNWFSAGSNISGVYKAGTVAAGADAKTIWVSP--EEGANAAISTDNG 585
Qy 604 NSWAAASQGVANQIRSDRNPNTFYALSNGTYRSTDDGVTF-QPVAAGLPSGGAGVGM 662
Db 586 NKWTPCSGLPGAKVRSRDNPKKFGYFLNGKFYSTDAGATFTQSSOTGLPTKGK-GI- 643
Qy 663 PHAVPGKEGLWLA-ASSGLYHSTNGGSSAITGVSAVNVGFGKSPAGSSYPAFVVG 721
Db 644 FKTVIGHEGDIWTAGGKDLWHSHTDSCATFTKVSVDASDVTGVLGSKTDDGYPAIYMDA 703
Qy 722 TIGVTCAYRSDCGTWTWLINDDOHQYGNMGQAITGDHANLRVYIGTNGRIGVYDGI 781
Db 704 TIDGTAGIFRSDDEGATWVRINDDAHQYSPDYCITGDPNKGVRVGTNGRIGVYDGI 763
Qy 782 GAPSPSPSPSVS 793
Db 764 GS-QPTPTPSVT 774

RESULT 4
Q9P4T8 PRELIMINARY; PRT; 806 AA.
AC Q9P4T8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE CEL6 PROTEIN PRECURSOR.
GN CEL6.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
```



```
Db 472 DLDRAPOAHYHTPYGTTNGIDYAGNKPNSINVRSGAD--DYPT--LALSSNFGSTWYA 526
Qy 564 GSEPGGVTTGTTVAASADGSRFVWAPDGPQVPVYAVGFGNSWAASOGVPANAQIRSDRV 623
Db 527 DYAASTSTGTGCAVALSADGDTVLLMSSTSGALVSKSQG---TLTAVSSLSFGAVIASDKS 583
Qy 624 NPKTFYALNGTFRYRTDGGVTQPVQVYAGLPSGAGVGMFHAVPGKGGDLWLAAASSGLYH 683
Db 584 DNTVFYGGAGAIYVSKNTATSFKTYS-LGSSSTTVNAI-RAHPFSIAGDVYASTDKGLWH 641
Qy 684 STNGSSWSAI-TGVSAVNVGFGKSAPGSSYPAFVVGVTGGVYAGYRSDDCGTTWVLI 742
Db 642 STDVGSFTIGSGVTAGWFGFGKASSTGSYVYIGFFIDGAAGLFKSEADAGTNQVVI 701
Qy 743 NDDHQYGNMQA--ITGDHANLRVYIGTNGRIVYDGGAPSG-----SPSPSV 792
Db 702 SDASHGFGS-GSANVNVGDLQTYGRVGRGHERPGHLLRSQREPAGRHGDDGDTTTSKT 760
Qy 793 SPSPASLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852
Db 761 STVSTTLKTTSSASTSSSTVTKTTSSSTTSKASSTTTTNTTTTSTSSGTTATA 820
Qy 853 SS 854
Db 821 SA 822

RESULT 6
ID Q9WE1 PRELIMINARY; PRT; 707 AA.
AC Q9WE1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOGLUCANASE, PUTATIVE.
GN TM0305.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001712; AAD35393.1;
DR TIGR; TM0305;
DR InterPro; IPR002860; BNR.
DR Pfam; PF02012; BNR; 9.
KW Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743C8DA6A99F CRC64;
```

Query Match 22.6%; Score 1161; DB 16; Length 707;
Best Local Similarity 34.6%; Pred. No. 7.2e-51;
Matches 267; Conservative 126; Mismatches 280; Indels 98; Gaps 27;

```
Qy 32 LGVLPIAITASPAHAANTQPTWNSVNAIGGGFVDGVFNEGAPGILYVTRDIGMYRWD 91
Db 10 LAIGLVVFGAT-----FEWKSVEINGGVFPGIIFHPASPGLLYARTDVGGLYRWD 60
Qy 92 AANGRWIPLLDWGNWNGYNGVVSIAADPTNTNKVAAVGMVNSWDPNDGAILRSDQ 151
Db 61 EETRWKQLDFLRDQSDYNGVLSVALDSPDKRIYATMGKYTDQW-AGYGAILISEDY 119
```

```
Qy 152 GATWQITPLP---FKLGGNMPGRGMRGLAYDPNNDNILYFGAPSGKGLHRSTDGATWS 208
Db 120 GETTIVNLNDKYGIKGGNEDGRNAGERLQVDFNFSVLFEMGT--TKYGLKSKSEDFGNWK 178
Qy 209 QMTNFPDVGVIYANPTDTCYQSDIQGVVWAFDKSSSSSLGQASKTIFVGVADPNPNVFW 268
Db 179 KVDSPFST-----SVTFVLFDEKSGEKSPTPRIFVGCSEPKG-IFV 219
Qy 269 SRDGGATQWAVPGAPTGFIPHKGVDFPNVHVLVIATSNGTGGPYDSSGSSGVKFSVTSGTW 328
Db 220 TEDGGTTNVNLPNPLNDIPLRGKIH--DGILYVTLNAGLPGNGATRGAVKVKYVIADQKW 277
Qy 329 TRISPPVSTDTANDYFGYSLGTLTDROHPNIMVATQISWHPDIIIPRSTOGGATWTIWD 388
Db 278 YDTPMKGD-----FGYCGIDVQE---NVVISTLDRWYPHDEIFISLNGTWTWRPLLE 328
Qy 389 WTSYPNRSRLRVLDISAEPWLTFCGVQNPVPPSPKLGWMDMAIDPFNSDRMLYGTGAT 448
Db 329 KANF-----DINKAPWK---DLNP-----HWISD-VKIDPFDMNRAIFTTGYG 368
Qy 449 LYATNDLTWKDSGGQIHIAPMVK-----GLEETAVNDLISPPGAGPLISALDGLGFTH 502
Db 369 VWVYELKKSPEG---MGRPVKWIIFENRGLLEETVVLQLPPIGERPLLSAIAIDWGGFRH 424
Qy 503 ADTAVPSTIFTSPVFTTGTSDYAEINPSIIVRAGSFDPSOPNDRHRAVAFSTDGKNWF 562
Db 425 ESLDTPSSMI-KPLKWTSLGIAFAYONSKFVARVHTY---TYP---FLSIEDGGINWR 477
Qy 563 Q-GSEPGGVTTGG--TVAASADGSRFVWAPDGPQVPVYAVGFGNSWAASOGVPA---NA 616
Db 478 EIEVPEGITDGGRLSLAVSNDGKTLWSPAN--HEVIVSSDKGSKWKKAISVPVPEFY 535
Qy 617 QIRSDRVNPKTFYAL--SNGTFFRSTGGVTQPVQVYAGLPSGAGVGMFHA----VPGKE 670
Db 536 FPASDPVNPSEKFIYDFWKNGDFLISKDGKSEMK-GAKLPSEFDMWYSLSFYPVLAADRE 594
Qy 671 GDLWLAAS-SGLYHSTNGSGSSSAITGVSSAVNVGFGKSAPGSSYPAFVVGVTIGGYTGA 729
Db 595 GDIWLALQWNGLYRSKDGGITFERLGNVDIAYVIGFAGAPKPGTDPYPAIYLNGWVNGYGI 654
Qy 730 YRSDCCGTTWVLINDDQHQYGNMQAITGDHANLRVYIGTNGRIVYIGDI 780
Db 655 FMSTDECKTWMRINNDKHQFG-WIHYMIGDMNERGIFLGTGEGRIIVGEV 704

RESULT 7
ID O52374 PRELIMINARY; PRT; 1779 AA.
AC O52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.8).
GN XYNC.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1.
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036924; AAB95326.1;
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBD_3; 3.
DR Pfam; PF02018; CBD_6; 2.
DR Pfam; PF00331; Glyco_hydro_10; 1.
```


[illegible]

```

Query Match      8.0%; Score 409; DB 2; Length 1770;
Best Local Similarity 20.0%; Pred. No. 1.3e-12;
Matches 263; Conservative 163; Mismatches 386; Indels 504; Gaps 58;

Qy 14 RRLVSLLAATASFVAAALGVLPITAITASPAHAAPTQPTWTNSNVAIGGGFVDGVTFNEG 73
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KRVTISL-SLLEFFLINTLVGLT--IFHOAKAAA--YT-----VDLEG 40

```



```
RESULT 10
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)
DE (CELLULOSE A) (FRAGMENT).
GN CELA.
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Anaerocellum group; Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A..
RC STRAIN=2-1320;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ENDORHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY E
CC (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULOSE FAMILY L
CC (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC EMBL; Z86105; CAB06786.1; -
CC HSP; P26221; ITF4.
CC InterPro: IPR001956; CBD_3.
CC InterPro: IPR000556; Glyco_hydro_48.
CC InterPro: IPR001701; Glyco_hydro_9.
CC Pfam; PF00942; CBD_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC PRINTS; P00844; GLHYDRASE48.
CC ProDom; PD001947; CBD_3; 2.
CC ProDom; PD011903; Glyco_hydro_48; 1.
CC PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT NON_TER 1
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 7.8%; Score 398.5; DB 2; Length 1711;
Best Local Similarity 22.9%; Pred. No. 4.3e-12;
Matches 197; Conservative 116; Mismatches 267; Indels 281; Gaps 40;

QY 262 PNPVFNRSRGATWQA-----VPCAPGTFPHKGVDVFNHVLVIATSTNGGPDGS--S 315
DB 25 PN-----WVRN---NWRGDSALKDQDGLDITGWFADGHDVRF-----NLPMSYTGFTLS 73
QY 316 GDVWKFS---VTSGTWTFRISVPSTDFANDYFGVSGULTIDROHPTIMVATQI----- 365
DB 74 WAVEYKDAFVKSQLEHI--LNQIEWNDYF-----VKHPKSVKYYVQVGDGSKDH 124
QY 366 SWW-----PDTIFRSTGGATWTRIDWTSYPNRSLRYLDISAEPMLTFGVQP 415
DB 125 AWMGPAEVMQMERPSFKVTQSSPGSTVVTE---TAASLAAASIVLK-----DR 169
QY 416 NPPVPSPKLGWDEMAIDFNSDRMLYGCATLYATNDLTKWDGSGQIHIAWPVGLLEE 475
DB 170 NPTKAATYLLQHAKE-----LYEFAEVTKSDAG-----Y 197
QY 476 TAVNDLISPPSG-----APLISALDGLGGFTHADVTAVSTFTSTSPVFTGTSVDY 526
DB 198 TAANGYNSGSGFDELSWAAMVLYLATNDSTYLTKE-----SYVONWPKISGNTIDY 252
```

```
QY 527 -----AELNPSTIIVRAGSFDPSQPNDRHVAFTDG-----GKNWFQ 563
DB 253 KWAHCWDDVHNGAAL---LLAKITGKDIYKQIITESHLDIYWTGYNGERIKYTPKGLAWLD 309
QY 564 --GSEPGGVTTGGTVAASAD-----GSRFVWAPGDPGPVYVAVGFG-- 603
DB 310 QWGSRLRYATTATFLAFVYSDWVGPCPKETRYKRFGEISQIDYALGSAGRS--FVVGFGTN 367
QY 604 -----NSWAASQGVPA-----NAQIRSD--- 621
DB 368 PPKRPHRHTAHSSWADSQSPSYHRHRTLYGALVGGPGSDSYTDDISNYVNVNACDYNA 427
QY 622 -----RVNP-----KTFYALNSGTFFYRSTDDGGVTFQFVAAGLPSSGAVGVNFH 664
DB 428 GFVCALAKMYQLXGGNPIDPFKALETNTDEFF-----VEAGINASGTNFEIK 476
QY 665 AVPGKEGDLWLAASS-----GLYHSTNGGSSWSAITGVSSAVNVWGKSNAP----- 710
DB 477 AIVNNQSG-WPAKATDKLKFYFVDLSELIKAGYSPNQLT--LSTNYNOGAKVSGPYWMDA 534
QY 711 --GSSYPAVFVVGTI--GGVTGAYRSDCGTTWVYLINDQHQYGNWGOAITGDHANLRRV 766
DB 535 SKNIYYILVDFTGTLIYPGGQDKYKE-----VQFRIAPQNVQWVNDNSNDYSFODIKGV 588
QY 767 YIGT-----NGRGIYVGDIGGAPSGSPSPSVSPASPSLSLSPSPSPSPSPSPSPSP 816
DB 589 SSGSVKTKIPLYDGDVKVMDGPGGSGATPTTA-----TATPTTPTVTPPTPTPTP 642
QY 817 SSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 875
DB 643 TS--TATPTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 690
QY 876 QYKNDSAPGDNQIKPGLQVNVNTGSSSDVLSVTVRVWFTRDGGSSFLVNCDDAAIGCG 935
DB 691 LYANKETNSTNTIRPWLKVNTGSSSIDLSRVTIRVWTVVDGDKAQSALS-DWAQIGAS 749
QY 936 NIRASFGSVNPATPTADTYLQ 956
DB 750 NVTFKFKVLSVSGADYYLE 770

RESULT 11
Q39620 PRELIMINARY; PRT; 473 AA.
AC Q39620;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VSP-3 PROTEIN PRECURSOR.
GN VSP-3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC-621;
RC MEDLINE=95093034; PubMed=8000007;
RA Woessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,
RA Goodenough G.W., Haring M.A.;
RA "Domain conservation in several volvoclean cell wall proteins.";
RL Plant Mol. Biol. 26:947-960(1994).
DR EMBL; L29029; AAB53953.1; -
KW Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 473 POTENTIAL.
SQ SEQUENCE 473 AA; 47532 MW; 38506131FAA674A4 CRC64;

Query Match 7.3%; Score 372.5; DB 10; Length 473;
Best Local Similarity 33.0%; Pred. No. 2e-11;
Matches 132; Conservative 36; Mismatches 139; Indels 93; Gaps 17;
```



```
QY 492 SALDGLGGFTHADYVAPSTIFTSPVTTGTSVDYAEALNPSIIYRAGSFDPSSQPN---- 547
Db 30 SAAGPYTYTITANSQVIATSLRLSLV---ETNLKY--LTPGGLARKNPITQTSQPSAAQL 84
QY 548 ---DRHVAFSDDGKNW---FQGGPEPGVTTGGTVAASADGSRFWAPDGPQPVVYAVG 601
Db 85 AGADAYVILSNRYISYWDTERKMGSPKNV-----RDLKDVNAGGS-----LVLLDG 131
QY 602 FGNSWAASQGVPAANAQIRSDRNVKPTFYALSNG---TYRSTPDGVTQFPVAAGLPSSGAV 659
Db 132 YSTASGTNTFVLQIDAVLGTAKSGCTGALYNGVNVYRRANSSSPFGKITPSLLVKGGS- 190
QY 660 GVMFHVPGKECDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVFGKSA---PGSSYPA 716
Db 191 -----SRGESGLTGTSGAVLFSSNPDKMTKAVT--ASAITWSVGKGAITFGSSFAM 241
QY 717 VFVVGITGGVTGAYRSDDCGTTWVLINDQHOYGNWGOAITGDHANLRRVYIGTNGRGIV 776
Db 242 PHLKG-----YEDMS-----GVAVT--LANENRT----- 263
QY 777 YGDIGGAPS-----GSPSPSVSPASPSLSPSPSSPSPSPSPSPSPSPSPSPSPSPSP 830
Db 264 ---GASPSPKASPSPKVPSPKASPSPSPKASPSPSPKASPSPSPKASPSPSPKASPSP 319
QY 831 SPSPSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
Db 320 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 359

RESULT 12
Q92411
ID Q92411 PRELIMINARY; PRT; 997 AA.
AC Q92411;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CELLULOSE PRECURSOR (EC 3.2.1.4).
GN CELB
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1tf4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000566; Lipocin_cytrFABP.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 POTENTIAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;
```

Query Match 6.8%; Score 350.5; DB 2; Length 997;
Best Local Similarity 21.4%; Pred. No. 5.9e-10;

RESULT 13
Q44562

```
Matches 214; Conservative 109; Mismatches 294; Indels 383; Gaps 47;
QY 125 NKWAAAYGMYTNS---WDPNDGAILRSSDQATWQITLPLFKLGNMPPGMRGLAVDP 181
Db 147 NELWQGVQAGNTDHAMGPAAE-----VMQNRPSFKIDASCPSGLAAATAAL 195
QY 182 NNDHILFVGAPSGKGLWRSDSGATWS-----QMTNEPDV--GTYIANPTDT----- 226
Db 196 AASIVF-----ADSDPVYSAKLQHAKEYLNFAITYRGKYTCITDAAAFYN 243
QY 227 --TCYQSDIO-GVVWV-----AFDKSSSSILGOASKTIFGVGVADPNPNVFW 268
Db 244 SWTGYDELAGGANWILYATNDNAYLSKALSAADRWTSGSA-----NWYTW 292
QY 269 SRDGGATWQAVPGAPTGFIPHKGVDFPNVHLYIATNTGPGYDGGSSDVKWKFVSTGTW 328
Db 293 TQG---WDS-----KHYG---AQILLARITSNLNP-----EATKF----- 322
QY 329 TRISPVSTDTANDYF--GYSGLTIDRQHPTNINVAQISHWPTIIFRSDGDAWTTRI 386
Db 323 -----IOSTERNLDYWTGTNGGRV-----KTPGGGLAWLDQ 354
QY 387 W-----DWTSPNRSRLRYVLDISABPWLTFGVQPNPVPSPKLG--- 425
Db 355 WGSRLYAANAAFISFYSDWVSDVPKKSRY--QNFATSOINVLGDNPROSSYVVGVCQN 412
QY 426 -----WMDEAMAIDPFNSDRMLYGT-----GATLYATNDL-----TKW 458
Db 413 SPQHPHRTAHSSWMNEDI--PANHRILYGAVMGPNASDQYTDGIDGYVSNEVATDY 470
QY 459 DSG-----GQIH-----IAPWKGLE-----ETAVNDLISPPSG 487
Db 471 NAGFTGALAKMNLLYGQNHQPLANFPAPKVGDEYFVEAAVRRSSGSNYTEIRALLNRS 530
QY 488 APLISALDGLGGFTH---ADVTAVPSTIFTSPVTTGTSVDYAEALNPSIIRVAG----- 538
Db 531 WP--ARMGDQLSFYFLDLSEVYAGRTV--SDVQTVSSSEGATVQPVVWDAKRIYA 586
QY 539 -----SEDPSPQPNDR-HVAFSTDGKN-WFGGSEPG--GVTTGG----- 574
Db 587 ITANFSNTKIYPGGEGNYRKEVQFRITGPGQANPANDPSYQNTLTGNPNVKSNIPIVYDA 646
QY 575 --TVAASADGSRFYWAPDCOPVYVAVGFGNS---VQAVA-GNSQVALNWSAGAVSYTVKRAEVS 626
Db 647 GKVSGGEPGVTVPVAVPAAPAG--VQAVA-GNSQVALNWSAGAVSYTVKRAEVS 703
QY 627 TFYALS-NGTFYRST--DGGVTFQPVAAGL-----PSSGAVGVMFHVPKGEGLWLAA 677
Db 704 TTVAGVNGLTYYTNTGLTKNGKTYVVYVTVNSAGESPASVQVSGMPQAATVPQAVTL 763
QY 678 SSGLYHSTNGGSSWSAITGVSSAVNVGFGKSPGSSYPVAVFVGTIGVGTGAYRSDCGT 737
Db 764 TAG---NQNNMLSWTAASGASYT-----VQRAVAG--GTYTDAVATGL 801
QY 738 TWVLINDQHOYGNWGOAITGDHANLRRVYIGTNGRGIVYDIDGAGSPSPSPSPSPAS 797
Db 802 AVLNYNDT-----TALNGTYSYRVAAVNASGQTL----- 831
QY 798 PSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 857
Db 832 ----- 832
QY 858 SSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 916
Db 833 NVVMLTPSGPPVSTGTLEVOYRSGSGNSSNAVTPQFNKLNKTGTQAIDLSTVKIRYFTK 892
QY 917 DGGSTLVYNCDAWAAIGCGNIRASFGSVNPAATPTADTYLQ 956
Db 893 D-GTEELSFWDYDAQVGSANVOGMFVAVNPAKGTADTYVE 931
```


Db 241 TSGFIDDLGW---AAVWLXIATDSSSYLTKAEELMSEYANGTNTWTQCWDDVRYGTLIM 297
QY 100 LLDVWGNWNGYNGWSIAADPINTNFWAAVGM-YTNWSDPNDGAILRSSDQG----- 152
Db 298 LAKITGREL--YKCAVERNLDHWDTRITYPKGNAYLTGW---GSLRYATTAAFLACVY 351
QY 153 ATWQ-----ITPLPKLGNMPGR-----GMGERLAVDPNNNNILYFGAPSG 194
Db 352 ADWSGCDNKKTKYLNFAKSIDYALGST--GRSFVVGFGTNYPOHPHRR----- 400
QY 195 KGLWRSTDGATWSOMTFDP-----VGYIANPTDTTGYOSDIQGVW--VAFDKSSS 246
Db 401 -----AHSWANSWKIPEYHHILYALVGGPGSDSDNDITDITVONEVACDYNAG 452
QY 247 SLGQAKTIFGVADPNPNFWSRGGATWQAVPGAPTGFPHKGVDPVNVHVLVIAT-- 304
Db 453 IVGALAK-----MYQLYGEPID--DFKAIETPTNDEIFVESKF 489
QY 305 SNTGGPYDGGSDGVWKFVTSVGTWTRISVPSTDTAN-DYF-----GYSGLTIDRQ 354
Db 490 GNSQGP---NYTEVISIYINRTGW---PPRVTDKLSFKYFIDLTETLIQAGYS----- 535
QY 355 HPNTIMVATOISWPPDTIIPRSTGGATWTRWDWTSYPNRSRLAYL-DISAEPLWTFGV 413
Db 536 -PDVVKVDT-----YIEGKLSGPYVWD-----KNRIYIYLVDSGTK-----I 575
QY 414 QPNPVPSPKLGW---MDEAMAIIDPNSDRMLYGTGATLYATNDLKWDSGGQIHIAPM 469
Db 576 YPGGEVEHKKQAQKISVPGQYWDPTN-DPSYKGLTSQLEKNKYIAAYDNN-----L 628
QY 470 VKGLEEAVNDLISP-----PSGAPLISALGDLGTFHADVTAVSTFT-----SPVFT 519
Db 629 VWGLEPGAATSTPATPTPTPTPTPTVTA-----TPTPTPTPTGSP--G 674
QY 520 TGTSDVYAEINPSIIVRAGSFPDSSOPNDRHVAFTSDGKNWFOGSEPGVTTGCTVAAS 579
Db 675 TSGVKVLYKNNETSASTGSIKP-----WFK-----IVNGG--SSS 708
QY 580 ADGSRF---VW--APGDPGQPVVAVGFGNSWAASQGVPAANAQIRSDRVNPKTFYALNSG 634
Db 709 VDLRSVRKIRYWTYVDGKPOSVC-----DW-----AQIGASNVT--FNEVKLSSG 752
QY 635 TFYRSTGGVTFQVAAGLPSGAGVGNFHAVPGKE--GDLWLAASSGLYHSTNGSSWSA 693
Db 753 V-----SGADIY--LEVGF--SSGAGQLQ-----PKDGTGDIQVRFNKNDWSYNQADWSW 800
QY 694 ITGVSSAVNVGFGKSAFGSSYPVAVVVGTTGGVTGAYRSDDCGTWVLINDDQHQYGNWG 753
Db 801 LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG 824
QY 754 QAITGDHANLRVYIGTNGRIGVYIGDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPS 813
Db 825 Q-----EPGGA-----TPAPTSTATPT 841
QY 814 PSPSSSPS 873
Db 842 TPTATPT 887
QY 874 KVQKNNDSAPGDNQIKPGLQVWNTGSSVDLSVTVRYWFTRDGGSSTLYNCDWAAIG 933
Db 888 KWWANGNLSPNTVLPKIKIENVGTAVDLSRVKRYWYTIIDGATQSV----- 938
QY 934 CGNTRASFSGSNPA 947
Db 939 -----SVASSINPA 947

RESULT 15

Q9RFX5

ID Q9RFX5

AC Q9RFX5;

DT 01-MAY-2000 (Tremblrel. 13, Created)

PRT; 930 AA.

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
MULTIDOMAIN BETA-1.4-MANNANASE PRECURSOR.
GN MANA.
OS Caldicoccus cellulosivorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10633733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicoccus cellulosivorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL: AF163837; AAF22274.1; .
DR HSP: O06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR004302; Chitin_bind_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF00942; CBD_3; 2.
DR Pfam: PF00150; cellulase; 1.
DR PRINTS: PR01217; PRICHEXTENSIN.
DR PRODOM: PD001947; CBD_3; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL.
FT SIGNAL 34 930 MULTIDOMAIN BETA-1.4-MANNANASE.
FT SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;
QY 560 NWF-----OGSE-----PGVTTGGTVAASADG---SRFW--APGDPGPVYAVGFG 603
Db 81 NWFGLNISAGRRHREIIPDGKLGPT--ASFDMNQARTDMWTRLPQCATITVR--- 135
QY 604 NSWAASQGVPAANAQIRSDRVNPKTFYALNSCTGYRSTGGVTFQVAAGLPSSGAVGVMF 663
Db 136 NAWAP-----HPGTWY-----LYTRDGDWPTQL----- 160
QY 664 HAVPGKEEDLWLAASSGLYHSTNGSSWSA--TGVSANVNGFGKSAF--GSSYPVAVFW 720
Db 161 -----KWSLEPTTFPFSQVTPNPPINSSGPDGAEY----- 188
QY 721 GTGGVTCAYRSDDCGTW-VLINDDQHQ---YGNWQAITGD-HANLRVYIGTNGRGI 775
Db 189 -----SWQVLPKQGRHIIYIMWRSDSPEAFYNCSDYFSGSPIAY 231
QY 776 VYGD--IGGAPSGSPS 833
Db 232 EFGDPRGGT-----MITP-----PPSGTPTPTPTPTPTPTPTPTPTPTPTPT 273
QY 834 PSRSPS 893
Db 274 P-----IVTPTSTPTPSA-----SGLRVEYRVGDTSATDNQMKPQL 310
QY 894 QVWNTGSSVDLSVTVRYWFTRDGGSSTLYNCDWAAIGCGNTRASFSGSNPATPTADT 953
Db 311 RIVNTGSAVPLTELKVRWYTKN--STQAEQYFCDWQAIGCSNIRAOFSVLSQPVSGADS 369
QY 954 YLQ 956
Db 370 YIE 372

Query Match 6.3%; Score 325.5; DB 2; Length 930;

Best Local Similarity 25.8%; Pred. No. 9, 9e-09;

Matches 109; Conservative 59; Mismatches 98; Indels 157; Gaps 21;

Search completed: July 2, 2002, 09:17:22

Job time: 330 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:34 ; Search time 3473.69 seconds
(without alignments)
17283.705 Million cell updates/sec

Title: US-09-917-376-2
Perfect score: 2869
Sequence: 1 atgcatcgttcggaacat.....ggcggacacctacctgcagn 2869

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

C	1	748	26.1	41906	1	SC5C7	AL031515 Streptomy
	2	512	17.8	3262	1	AF078038S1	AF078038 Caldicell
	3	402	14.0	2745	8	AB015511	AB015511 Aspergill
	4	375.6	13.1	12732	1	AE007608	AE007608 Clostridi
	5	327.6	11.4	2579	8	AB1292929	AB1292929 Agaricus
	6	131.6	4.6	3237	1	AF200304	AF200304 Caldibaci
	7	103.6	3.6	749	14	NFORP2	D13755 Multicapsid
	8	103.6	3.6	131995	14	OPU75930	U75930 Orgyia pseu
	9	100.8	3.5	390	6	AR135147	AR135147 Sequence
	10	96.4	3.4	4567	1	AF163837	AF163837 Caldibaci
	11	91.4	3.2	183665	2	AP003605	AP003605 Mus muscu
	12	90.2	3.1	40340	1	AP003605	AP003605 Nostoc sp
	13	86.4	3.0	183544	2	AC095146	AC095146 Rattus no
	14	84.2	2.9	97225	2	AC099417	AC099417 Rattus no
	15	83.6	2.9	118584	14	AY043265	AY043265 Epiphyas
	16	82.2	2.9	180104	9	DJ526N18	AF123462 Homo sapi
	17	81.6	2.8	13669	14	TTV1TP	X14855 Thermoprote
	18	81.6	2.8	99427	2	AP003821	AP003821 Oryza sat
	19	81.2	2.8	216819	2	AL645725	AL645725 Mus muscu
	20	80	2.8	5437	1	CASR69XYN2	AF036924 Caldicell
	21	80	2.8	11707	1	AF078737	AF078737 Caldicell
	22	80	2.8	11707	6	E35099	E35099 Truncated c
	23	79	2.8	15348	1	AE007163	AE007163 Mycobacte
	24	78.4	2.7	159974	2	AC021474	AC021474 Homo sapi
	25	78.2	2.7	77457	1	AF210249	AF210249 Streptomy
	26	78	2.7	1256	7	AF069776	AF069776 Xanthomon
	27	77.6	2.7	16671	1	AE006990	AE006990 Mycobacte
	28	77.6	2.7	67200	1	MTV017	AL021897 Mycobacte
	29	77.4	2.7	125020	9	AF429315	AF429315 Homo sapi
	30	77.2	2.7	63621	9	AL590867	AL590867 Human DNA
	31	76.8	2.7	4034	14	TTV1VP	X14717 Thermoprote
	32	76.8	2.7	165742	2	AC020697	AC020697 Homo sapi
	33	76	2.6	169915	2	AC068801	AC068801 Homo sapi
	34	76	2.6	173434	2	AC016142	AC016142 Homo sapi
	35	76	2.6	176463	2	AC012149	AC012149 Homo sapi
	36	75.8	2.6	13278	1	AE001712	AE001712 Thermotog
	37	75.6	2.6	203299	2	AP001178	AP001178 Homo sapi
	38	75	2.6	5513	1	AT286105	Z86105 A. thermophi
	39	74.8	2.6	47852	1	MTV023	AL022022 Mycobacte
	40	74.4	2.6	1171	14	BPU10884	U10884 Bacterioph
	41	74.4	2.6	15311	1	AE007164	AE007164 Mycobacte
	42	74.2	2.6	161355	2	AC011154	AC011154 Homo sapi
	43	74.2	2.6	192931	2	AC094068	AC094068 Rattus no
	44	74	2.6	11186	1	AE003939	AE003939 Xylella f
	45	73.6	2.6	56414	1	MTV002	AL008967 Mycobacte

ALIGNMENTS

RESULT	1	SC5C7	41906 bp	DNA	linear	BCT 07-SEP-1998
LOCUS	SC5C7/c	Streptomyces coelicolor cosmid 5C7.				
DEFINITION	AL031515	Streptomyces coelicolor cosmid 5C7.				
ACCESSION	AL031515	Streptomyces coelicolor cosmid 5C7.				
VERSION	AL031515.1	GI:3559990				
KEYWORDS		ATP/GTP-binding protein; lyase; narG; narH; narJ; nitrate reductase; oxidoreductase; RNA polymerase sigma factor; secreted cellulase; uvrA-like protein.				
SOURCE		Streptomyces coelicolor A3(2).				
ORGANISM		Streptomyces coelicolor A3(2).				
REFERENCE		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
AUTHORS		1 (bases 1 to 41906)				
JOURNAL		Seeger,K.J. and Harris,D.				
REFERENCE		Unpublished				
AUTHORS		2 (bases 1 to 41906)				
JOURNAL		Parkhill,J., Barrell,B.G. and Rajandream,M.A.				
REFERENCE		Direct Submission				
TITLE		Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project,				
JOURNAL		Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge				
		CB10 1SA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Prof.				
		David A. Hopwood, [3] John Innes Centre, Norwich Research Park,				

Colney, Norwich, Norfolk NR4 7UH, UK
 3 (bases 1 to 41906)
REFERENCE
AUTHORS Redenbach,M., Kieser,H.M., Denapaita,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
COMMENT Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5C7 lies between 1E6 and 4B5 on the Aser-A genomic restriction fragment.
FEATURES
SOURCE
 1..41906
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 5C7"
 complement(1..136)
 /gene="SC5C7.01c"
 complement(<1..136)
 /gene="SC5C7.01c"
 /note="SC5C7.01c, unknown, partial CDS, len: >45 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SC5C7.01c"
 /protein_id="CAA20613.1"
 /db_xref="GI:3559991"
 /db_xref="SPTREMBL:O86698"
 /translation="MORGSDRMSVHRDDEMKHELOGLLRSGHPTRSEWNDEPAEAD D"
 complement(142..146)
 /note="possible RBS upstream of SC5C7.01c"
 336..339
 /note="possible RBS upstream of SC5C7.02"
 345..2735
 /gene="SC5C7.02"
 345..2735
 /gene="SC5C7.02"
 /note="SC5C7.02, uvrA-like protein, len: 796 aa; similar to Streptomyces peucetius TR:Q54827 (EMBL:L76359) daunorubicin resistance protein drrc (764 aa), fasta scores: opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity in 775 aa overlap, and to many uvrA proteins e.g. UVR_A_NEICO excinuclease ABC subunit A (950 aa), fasta

scores; opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in 953 aa overlap. Contains 2x PS00017 ATP/GTP-binding site motif A (P-loop) and 2x Pfam match to entry PF00005 ABC_tran, ABC transporters, scores 42.40, E-value 1.7e-10 and 71.20, E-value 2.2e-17"
 /codon_start=1
 /transl_table=11
 /product="uvrA-like protein"
 /protein_id="CAA20614.1"
 /db_xref="GI:3559992"
 /db_xref="SPTREMBL:O86699"
 /translation="MSSAKRPGTGPESHVADSHDLIRVHGARENKLDVSDIPKRR LVFTGCVSGSKSLVFNTIAAESQRLINETYSAFVGFMPTLARPEVDLGLTAT IVDQRMGADPRSTVGTATDVNAMRLILFSLRGEPRIGPPSAYSENTASVRASGATV ERGNKAVRATFERTGCMCTHCEGRGTVSDIDLTYDDSKSLAGGAFIPGWKSDSQ WYQVYQSGFVDPDKPIREYTEKELRDEPLYGEPVKVNGVNLVYGLIPKIQSEFL SXDKAMQPHIRAFVERAVFTTCCECEGTRLSEGRSKIKKISIAQACAMEIRDLA EWRDLTSPVAPLTALTALDTLDSEVIGLYLSLDRPAGTUSGGEAQKVMIRHLGS STDTYTFDEPTVGLHPIQRMNDLLRLKNGTVLVVHEKPEATAIADHVVDLG PGAGTAGTVCFEETVEELRAADVTGRLHDDRVLKESVRPAGALETRDARTNLQ GVDVDPGLCVKVTGAGSKSLIHGSPAGADVTVSDQSPKIGSRSSNPATVTGL LPIRKAFKANGVKPALFSANSEGACPTCNGAGVYTDLANMAGVATPCDECEGRF QPAYLEYRGGRISEVLAMSDQAEFFEGAGEARTPAHKILQRLSDVGLYLILGQ PLTUSGGERQRLKATHMGEGGVTVLDEPTTGLHLDVBEQLGLDLRLVQAGSVI VIEHHQAVNAHADMIIDLPGAGHDGGRVVFEGTADLVADRSTLTUTGEHLAAYGA 471..1700
 /gene="SC5C7.02"
 /note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 42.40, E-value 1.7e-10"
 492..5515
 /gene="SC5C7.02"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 1893..2627
 /gene="SC5C7.02"
 /note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 71.20, E-value 2.2e-17"
 1914..1937
 /gene="SC5C7.02"
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
 2803..3285
 /gene="SC5C7.03"
 2803..3285
 /gene="SC5C7.03"
 /note="SC5C7.03, probable transcriptional regulator, len: 160 aa; similar in N-terminus to many e.g. RAMA_ENTCL transcriptional activator rama (113 aa), fasta scores; opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96 aa overlap. C-terminal 50 aa are unique. Contains Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, arac family, score 71.40, E-value 1.9e-17"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="CAA20615.1"
 /db_xref="GI:3559993"
 /db_xref="SPTREMBL:O86700"
 /translation="MTTLEDLARLRRLDMREYAEPDLDPALARVALMSAGHSRS FRAAYGEPTYSYLMTTRIRRIERAMALLRRGDMSTVCFAVGCTSLGFSRSRFTLVGES PSAYRERDHDCAALPACVAKVYVTRVNRGEPGRSGTPAGKSESVRNGEAAPAPPVA "
 2872..3132
 /gene="SC5C7.03"
 /note="Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, arac family, score 71.40, E-value 1.9e-17"
 3291..3707
 /gene="SC5C7.04"
 3291..3707
 /gene="SC5C7.04"
 /note="SC5C7.04, possible lyase, similar to e.g. LGUL_HAEIN lactoylglutathione lyase (EC 4.4.1.5) (135 aa), fasta scores; opt: 124 z-score: 215.2 E(): 9.7e-05, 29.0%


```

identity in 131 aa overlap"
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAA20616.1"
/db_xref="GI:3559994"
/db_xref="SPTREMBL:O86701"
/translation="MDISLSOCFTAVDDHDFALAFYRDVLGLEVRNDVGVEGMRWVTL
GSPAQVDIVLEPLDPAKASADKQAMADLLAKLHGLRGIILATDDVDATFEEHRAA
GAELVQEPDQPYGVCDFRPAKGNMVRHLRPRKG"
3978..3981
/note="possible RBS upstream of SC5C7.05"
3994..4794
/gene="SC5C7.05"
3994..4794
/gene="SC5C7.05"
/note="SC5C7.05, probable RNA polymerase sigma factor,
len: 266 aa; similar to S. coelicolor sigma factors of the
sigB subfamily e.g. TR:052313 (EMBL:AF036131) putative RNA
polymerase secondary sigma factor sigG (263 aa), fasta
scores: opt: 1258 z-score: 1540.5 E(): 0. 72.9% identity
in 251 aa overlap, and to RPOF_STRAU RNA polymerase
sigma-F factor (297 aa), fasta scores: opt: 713 z-score:
858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible
alternative start 216 bp upstream, not supported by
FramePlot. Contains helix-turn-helix motif at aa 228-249
(score 2104, +6.35 SD) and Pfam match to entry PF00140
(sigma70, Sigma-70 factors, score 210.40, E-value 2.8e-59"
/codon_start=1
/transl_table=11
/product="putative RNA polymerase sigma factor"
/protein_id="CAA20617.1"
/db_xref="GI:3559995"

```

Query Match	26.1%	Score 748	DB 1	Length 41906	
Best Local Similarity	59.8%	Pred. No. 1.4e-74			
Matches 1429	Conservative	0	Mismatches 895	Indels 64	Gaps 8
QY 154	tacacctggagcaacgtggcgatcgggggcgcgctttgtcgcagcgatcgtcttcaat	213			
Db 34351	TACACCTTGGAGAACGCCGCATCGACGGCGCGGTTTCGTCCCGGATCGTCTTCAAC	34292			
QY 214	gaagggtcacccgggaattctgtactgtcgagcgacatcgggggagtgtatcgtatggat	273			
Db 34291	CGGACCGAGAAGACCTTGGCTACGCCCGCACCGACATCGAGCGGCTACCGCTGGCAG	34232			
QY 274	gcgcgcaacggcggttgatctcttctggtatggtggatggaacaattgggggtac	333			
Db 34231	GAGGAGTCCACACACTGGAGCGCGCTCTCGACACACTCGCTGGGACGACTGGGGGCAC	34172			
QY 334	aacggcgtctgcagcatctgcggcagaccgatcaataactaacaaggatggcgccgcgtc	393			
Db 34171	ACGGCGTGTGTGCTCTGCGCTCGACGCGCTCGATCCGGACCGGGTCTACCGCGCGGTC	34112			
QY 394	ggaatgtacaccaacagctgggaccccaaacgacggagcgattctccgtctctgtatcag	453			
Db 34111	GGCAGGTACACCAACGACTGGGACCCGACCAACGGCGGGTCTCGCTCGCGCGACCGG	34052			
QY 454	ggcgcaacgtggcaaaataacgcccctcggttcaagcttggcggaacatgcccgggcgt	513			
Db 34051	GGTGCAGCTGGGAGAGGCGGACCTCCGCTTCAAGTGGCGGCAACATGCCCGGCGC	33992			
QY 514	ggaatggcgagcgcttgcggtggatcccaacaatgacaacattctgtatttcggcgcc	573			
Db 33991	GGCATGGGTGAGCGCCTTGGCGCTGCACCCGACGACACGAGTGTCTGTACCTGGGCGCC	33932			
QY 574	ccgagcgcaaaagggtctctggagaagcacagatcccgcgcgactggtcccgatgacg	633			
Db 33931	CCACGCGCCACGAGCTGTGGCGGTCCACCGACGCGGGCGTCACTTGTGCGAGGTGACG	33872			
QY 634	aactttccgagctaggcagctacattgcgaatccccactgacacgacggcgtactagagc	693			
Db 33871	GCCTTCCGAACCCCGGGAATCTACGCGAGGACCCGACACGACGCTCGGCTACGCCCTCC	33812			

Qy	694	gatatccaaggcgtctgttgggttcgcatttgacaaagtcttcgtcatcgctcggaagcgy	753
Db	33811	GACAAACCAGGGCATCACCTTGGGTTCACCTTCGACGAGTGCGACGGGCCGGCCGGCAGC	33752
Qy	754	agtaa--gaccatttttgtggqgtggcgatcccataatacccggtctcttggagcaga	810
Db	33751	GCCAGCGGCACCTCTACTCTCGGGTCTCGGACAGGAGAAGCGCGTCTACCGCTCGACC	33692
Qy	811	gacggcgcgagcgttgacaggcggtgccgggtgccccgcgcgcgcgttatccocgcacaag	870
Db	33691	CACCGGGCGCGCACTTGGGAGCGGCTCCCGGGACAGCGACGCGGTACTTGGCCCCACAAG	33632
Qy	871	ggcgctttgacctgctaaccacagtgctctatatgtccaccagcaaatcacgggtgtgctcg	930
Db	33631	GGTGTGCTGGACGGGAACCGGTACTGTACTTCGCTACAGGCACACCGCGCGCCCG	33572
Qy	931	tatgacgggagctccggcgactgtggaaaattctcgtgaacctcgggacatggacgcga	990
Db	33571	TACGACGGCGCGAAGCGCGCGGTGTATCCGGTTACGGACGGGCAACCGCACCTTGGACGGAC	33512
Qy	991	atcagccggtacttcgacgacacgccaacgactacttggtttacagcggtceact	1050
Db	33511	ATCAGCCCGCGCGGAGCGGACAC-----CTACTACGGGTCTCAGCGGCTTGACC	33461
Qy	1051	atcgaccgccagcaccgcgaacacataatggtggcaacccagatcatcgttgccgcgac	1110
Db	33460	GTCGACCGCAGCGCGCGGGACGGTGATGCGCACCGCGTAGACTCTGTTGGTGGCCGAC	33401
Qy	1111	accataatctttcggagcaccgcgcggctgtagcgtggaacgagatctggattggacy	1170
Db	33400	ACGCAGATTCTTCGGCTCGACGGACAGCGCGCGGACATGGTCCCAGCGCTGGAGCTTACACC	33341
Qy	1171	agttateccaatcgaagcttgcgatatgcttgacatttcggcggaqccttggctgacc	1230
Db	33340	TCGTACC CGGACCGGACAAACCGCTACACGATGACGTCTCTCTGTCCTCCCTGGCTCACC	33281
Qy	1231	ttcgggtctacgcgcgaatctctccgta cccagtcgcgaagctcggctggatggatgaagcy	1290
Db	33280	TGGSGCGGAACCCCGCGCCCGCAGCAGACCCCAAAGCTGGGTGGATGACGGAGGCT	33221
Qy	1291	atggcaatcgcgttcaaactctgtagctgtagtctctcggaaacaggcgcgcgttgtac	1350
Db	33220	CTGGAGATCGAACCGTTCGACTCGACCCGGATGATGTACGGCACGGGCGCCACCGGCTTAC	33161
Qy	1351	gcacaatatgatctcacgaagtggactccc---ggcgccagattcataatcgcgccgatg	1407
Db	33160	GGCAGGAGAACC TGACGAAC TTGGACACGAGGCGGCACGTTCTCGCGCTCGAGCCGATG	33101
Qy	1408	gtcaaaagattggagagacggcggf aaacgatctcatcagccgcgcgtctggcgcccg	1467
Db	33100	GTGCGGGGCCCTGGAGAGACGCGCGCTCAACACCTTGCGCTCCCCCGCTCGGGTGC CGCG	33041
Qy	1468	ctcatcagcgtctctggagactcggcggtctcaaccaacgcgcagcttactgcogtlgcca	1527
Db	33040	CTGCTACGCGCCCTGGGTGTAGCTCGCGCGCTTCGGGCACACCGAGCTGACCGAGGTCCCG	32981
Qy	1528	tcgacgatcttcacgttcaccgggtgttcacgaccggcaccagcgttcgactatgcggaa ttg	1587
Db	32980	TCGATGATGTACACCTCGCCCAACTTCACCTCGACGACGAGCGCTTGACTTCGCGGAGCG	32921
Qy	1588	aatccgtgcatacttcgcgc tggaa gtttcgatccatcgagccaacccgaacgacagg	1647
Db	32920	AAGCCGCGAGCTGTGTGTTCGGGCGGGCAATCTCGACTC-----CGGTCCG	32876
Qy	1648	cacgtcgcgttcttcgacagacggcggcgaagacttggttcccaaggcagcgaaactctggcggg	1707
Db	32875	CACATCGCGTTCATCCACGSAACAGCGGCCCACTGCTTCGGGGCACACCGACCCGTC GCGG	32816
Qy	1708	gtgacgacggcgcaacctcgcgcgaatcgggccacgagctctcgtttctcgtctgggctccc	1767
Db	32815	GTGAGCGCGCGGCACGCTCGCGCGGGCGCGACGCGACGCGCTTGCTGTGGAGCCCG	32756


```
Qy 2689 aataccgggtcgctgctgattgtcgacggtgcggtgactgttaccgcg 2748
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2878 AATGAGGCACACAGTGTGATCTTACGAGGCTTAAGATAAGTACTGTGTACACAGT 2937
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2749 gatggtgctcgtc 2762
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2938 GATGCTGACAGCC 2951
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AB015511
LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998
DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.
ACCESSION AB015511
VERSION 1
KEYWORDS Avicelase III.
SOURCE Aspergillus aculeatus cDNA to mRNA.
ORGANISM Aspergillus aculeatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Arai.M., Takada.G., Kawaguchi.T. and Sumitani.J.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Motoo Arai, Osaka Prefecture University,
College of Agriculture; Gakuencho 1-1, Sakai, Osaka 599-8531, Japan
(E-mail:motoo@biochem.osakafu-u.ac.jp, Tel:81-722-54-9465,
Fax:81-722-54-9465)
2 (bases 1 to 2745)
AUTHORS Arai.M., Takada.G., Kawaguchi.T. and Sumitani.J.
TITLE Avicelase III from Aspergillus aculeatus
JOURNAL Published Only in DataBase (1998) In press
FEATURES
source 1. .2745
/organism="Aspergillus aculeatus"
/db_xref="taxon:5053"
8. .2578
/gene="aviIII"
8. .2578
/gene="aviIII"
/codon_start=1
/product="Avicelase III"
/protein_id="BAA29031.1"
/db_xref="GI:3242655"
/translation="MARSSLLALCALGLADAAASQATWKNVVTGGGGFTPGIV
FNPASGVAVARTDIGAVRLNSDDTWTPLMDVWGNDDTHWDGIDALATDPVDTDRVY
VAYGMTNEDPNVGSILSTDOGDWTETKLPFKVGNMPPGRGMGERLAVDPNKSNI
LYGAKSGHLKWSSTDIYGATWSNVSFTWTGTYFQSSSTYSDPVGIAWTFDSTSG
SSGATPRIFPVGVADAGKSVKSEADAGATWAWVSGEPQYGFPHKGLVLSPEEKLTYIS
YANGAGPYDGTNGTIVHKNITSGWTDISPTSLASTYGYGLSLVDLPVGLTMAAL
NCWMPDELIFRSTDSGATWSPIEWENGYPSINYYSYDISNAPWIDTSTDOFPVVR
GMVEALADIPDSNHLVGTGLTVYGHDLTNWDSKHNVYKSLAVGLEMAVLGLI
TPGGPALLISAVDGGDFVHSDLDAAPNQAYHTPTVGTNGIDYAGNKNPSNIVRSAS
DDYPTLALSNGFSTWYADASTGTGAVALSDGDTVLLMSSTSGALVSKSQGL
TAVSSLPASVIAISKSDNMTFYGGSAGAIYVSKNTASFTKTVSLGSSSTVNAIRAH
PSTAGVDAKIDGLHSDTDYGTFTQIGSVTAGWGFSGFKASSTGVYVYIGFTFI
DGAAGLFHSDAGTANQVTSIDASHGSGSANVNGDLOTYGRVFRGHRPRLHROS
OREPACRHGDGDDTTSKTSSTVSTLTKTTSSASTSSSTVTKTTSSSSTSKAS
STTTTKTTTSTTSSTGTTATASAYACQCGNGHTGATVETGYTCTYSNRFVSOQVPS
"
BASE COUNT 571 a 803 c 820 g 551 t
ORIGIN

Query Match 14.0%; Score 402; DB 8; Length 2745;
Best Local Similarity 52.7%; Pred. No. 4.9e-36;
Matches 1211; Conservative 0; Mismatches 1020; Indels 67; Gaps 13;

Qy 89 ccgctctggaggtctgccatcgataacggcttctctgcgcgcgcgcgcgactc 148
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 CCACCTTGGCTTCTATGTCGGGCCCTCTCTGGCAGAGCTTGCAGAGCTGCGGCGT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 149 agccgtacacctggagcaacgctgcgacatcgggggcgcgctttgttcga---cgggatcg 205
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
Db 1140 GGATCCA---GGACACGACTTCACCGACCAATTCCTCGGTTCC--CGTCGGCTGGATGG 1193
Qy 1283 atgaagcagatggcaatcgatccgttcaactctgatcgtgatcgtctacggaaacagcgca 1342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1194 TTGAAGCCCTTGCCATTGACCAATTTGACTTCAACACCACTGGCTCTACGGCAGCGGCTGA 1253
Qy 1343 cgttgtagcaacaataatgatctcaagaagtgggactccgcgccagattcatatcgcgc 1402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1254 CAGTATATGGCGGACAGCACTTGACCAATTTGGGACTCGAAGACACAATGTGACGGTCAAGT 1313
Qy 1403 cgaatggctcaaggattggagagacgcggtaaacgatctcatcagccgcgcgtctggcg 1462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1314 CACTGGCGGTCGGTATTGAGGAGATGCGAGTCTCGACACTGATCACCGCCCGCGCGGAC 1373
Qy 1463 ccccgctcaatcagcgctctcggagacctcggggcttcaaccacgcgcgaogttactcgcg 1522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1374 CGCGCTCTGCTCGCTCGCTGTTGGCGACGACGCGGGTCTTACCATCGATCTGACGCAG 1433
Qy 1523 tgccatcgagatcttcacgtcacccggtgttcacgacgcgcgcacacgctgacatcgcg 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1434 CGCCAAACCAAGCGTACACACGCCGACCTACGCGCACGACCAACGGATCGACTACGCCG 1493
Qy 1583 aattgataccgtcgatcatcgttccgctgctgaaagtcttcgatccatcgagaccaacgcgaacg 1642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1494 GCAACAAGCCTTCCAACATCGTCCGGTCCGAGCTAGCGA-----CGACT 1538
Qy 1643 acaggcagctcgctctcgaagacgcgcgcgaagaactggttccaaagcgagcgaacatcg 1702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1539 ATCCGACCCCTGGCTCTGCTCGAATCTCGGATCGACTTGGTATGACAGACTACGCCGCT 1598
Qy 1703 gcggggtgagacgcgcgcgcacgcctgcgcacatcgccgacgcgcctctcgttctcgtgg 1762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1599 CAACGACACGGGACCGGACGGTGGCTCTTCCCGCGGATGGTGACACGGTGTCTTCA 1658
Qy 1763 ctcccgcgatcccggtcagcgtgtgtgacgcagtcggaatttggcaactccctgggctg 1822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1659 TGTCCAGCACCTCGCGCGCTGTGTGTGCGAAGTCGCAGGG-----CACCTTGACTG 1709
Qy 1823 cttcgcaaggtgttccgcgcgaatccccagatcccgctcagacgcgggtgaatccaaagactt 1882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1710 CGGTTCTCTCTCTCCCTCGCGGCGGTCTATCGCTCTCGAAGTCCGACAAACACGGTCT 1769
Qy 1883 tctatgcctatcaatggaacctctcatcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1770 TCTATGTGGATCCGGCGGCCAATCTACGTGTCGAAGAACACAGACCACTCTCTTC---- 1825
Qy 1943 cgttcgcgcgcggtcttccgagacgcggtgcgctcgtgtgatcttcaacgcgcgtgcctg 2002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 --ACCAAGACCGTGAGCTGGGGTCCAGCAGCACCGCTCAACGCCATCCGCGCATCCCT 1883
Qy 2003 gaaaagaaggcagatctgtgctcgatcagcgcgggcttttaacctcaaccaatggcg 2062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1884 CCATCGCGCGGATGTCTGGCGGTGCGACCGCAAGGGCCTCTGGCATTCACGAGATTACG 1943
Qy 2063 gcagcagttgctctgcaat---caccgcgctatcctcgcgcggtgaacgctggattgta 2119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1944 GTAGTACGTTACCGACAGATCGGACGCGGCTACGCGCGGCTGAGGCTTTCGGTTTGCA 2003
Qy 2120 agtctgcgcgcggtctcataccaccgcgcttcttctgctcgcgcacatcgagagcggtta 2179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2004 AGCGCTCTCCACGGGACGATACGTCGTGATCTACGGGTCTTCCACCATCGACGCTGCGC 2063
Qy 2180 cggggcgctaccgcctcgcagactgtgggacacactgggtactatcaatgataccagc 2239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2064 CGGGGCTGTTCAAGAGCGAGGACGACGACCAACTCTGGCAGGTGATCTCGGATCGCTCGC 2123
Qy 2240 accaatcaggaattgg---ggacaagaatcaccggtgacacgcgaatttaccgcgcgg 2296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2124 ACGGATTCGGGTCGGGTCCGCCAAGCTGGTGAACGCGACCTCTCAGACCTTACGGCCGCG 2183
Qy 2297 tgta-cataggacgaacgcgcgtgggaattgtatatacgggggacatttggtggcgccgtcc 2355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2184 TGTTCGTGGGCAAGACGGCGCGGGCATCTTCTACGGCAATCCACGGGAGCCTGCCG 2243
```

```
Qy 2356 ggatcgccgctctcgtcg 2373
    || | || | || |
Db 2244 GGCACCACGGCGACGGCG 2261

RESULT 4
AE007608 12732 bp DNA linear BCT 27-JUL-2001
LOCUS Clostridium acetobutylicum ATCC824 section 96 of 356 of the
DEFINITION complete genome.
ACCESSION AE007608 AE001437
VERSION AE007608.1 GI:15023819
KEYWORDS
SOURCE
ORGANISM
Clostridium acetobutylicum.
Clostridium acetobutylicum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 12732)
Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatsov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
Genome Sequence and Comparative Analysis of the Solvent-Producing
Bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
PUBMED 11466286
REFERENCE
2 (bases 1 to 12732)
Childress,D., Zeng,Q. and Smith,D.R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
Location/Qualifiers
source
1..12732
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
73..2592
/gene="CAC0919"
73..2592
/gene="CAC0919"
/codon_start=1
/transl_table=11
/product="Probably secreted sialidase; several ASP-boxes
and dockerin domain"
/protein_id="AAK78895.1"
/db_xref="GI:15023820"
/translation="MKRIVSVWAGLSIIFTTGFVTHISAANKAAASVOSQGYKWDN
AKIGAGGYPAVIFNKTEDLIYARTDMGGAYRWDRANNKWIPITDGFSDWMLGCE
IAIDPDIITNRVYIAAGLYTNDWDENAYILSSQDKGNTWKRYQLPFKVGGMFGRNG
ERLIQIDNPDKILYLGARSNGLWKSDEYQOTWSKVDNFPDPTDGYVDQPNQYETADRV
GVWVETFDPTGTGSGPTOMYGAADKTGNILYVTNDGKTSWAVKQPKGYLPHHG
ILASDGLXISYNTCGPSDQGVWYKNTKGEWNTITPPAVGDTKSGFGISVDA
QNFNVVATLNRWPDDEIYRSTADGKTWKPIWNGNICYNRTILGYNLDYSAQPLDW
QNTGTPPDLVLKLGMMWGLEIDPFNSDRMFYGTATLYGTDLTNDKGNVDISV
KANGIEEACAVNVWPTKQAOLLSAVDDCGFYHDDITKVPKMMFTPNFSATSIDY
AESVPNVVRVGVNDT'SKNOQDKCGISYDGGKWNFSAGSNISGVYKAGTVAAGADAK
TIVWSPPEGANAAYSTIDNGKWT'PCGLPOGAKVRSRVNPKFYGFLNGKFYSTDA
GATFTOSSOGLTPKKG IFKTVIGHGDIWAGKGLWHSIDSGATFKVSGVDAS
DTVGLSKSTDGYPALYMDATIDGTAGIFRSDDGATWVRINDDAHQYSGPYCITG
DPNKYGRFVGTNGIVGIDISQPTPTPSVTLGVDVNDGDIILQRQYISGKITFSSTK"
complement(2634..2978)
/gene="CAC0920"
complement(2634..2978)
/gene="CAC0920"
/codon_start=1
/transl_table=11
/product="Protein related to MIFH/DOPD protein family,
function in bacteria is unknown"
```


Db 250 AAAACAGAAAAGATTTAATATATATGCTCGTACAGATATGGAGGACATATAGATGGAC 309
QY 274 gccgcgaacggcggtggatccctctctctgtgattgggtggatggagaaattgggggtac 333
Db 310 AAGGCTAATAATAAAGTGGATTCCAATTTACTTGAT-----GGATTTTCTGATGGACAATG 363
QY 334 aacggcgctcgaacattcgccagacggcatcaataactaacaagtatatggccgcgcgtc 393
Db 364 CTTTGGCTGGGAAGATATAGCAACTGACCCCAATAGACACTAATCGTGTTTATATAGCAGCA 423
QY 394 ggaattacacacacagctgggaccacaacacgagggagcgtattctccgctcgtctgacag 453
Db 424 GGTTTGTACACAAATGATTGGCAAGATGAAATGCAATATATCTTAGTTCCACAGGATAAA 483
QY 454 ggcgcgaacgtggcaataacgcgccctgcgcgtctcaagcttggcggaacaatgcgcgggcgt 513
Db 484 GGTAAACACATGGAAGATACCAACATTCCTAATTTAAGGTAGGAGTAAATATGCGGGGAAGA 543
QY 514 ggaatggcgagcgcttcggtggatcccaacaataagacaacattctgtatttcgcgcgc 573
Db 544 AATATGGGAGAAAGACTTCAATATAGATCCTTAATGATGATAAAATACCTTTATCTTGGAGCT 603
QY 574 ccgaggggcaaaagggctggagaagcacagattccgggcgcgacctgggtcccgagatgacg 633
Db 604 AGAAGTGCTAATGGACTTTGGAAAAGTGAAGACTATGGACAAACATGGTCTAAGGTTGAT 663
QY 634 aactttccgagcattagcaacgacattgcaaatcccaactgacacgacgcgcgtatccagagc 693
Db 664 AATTTCCCTGATACAGCGGACTATGTACAAAGTCCC-----TCAAAATGAATATACAGCA 717
QY 694 gatattcaagcgctcgtcgtgggtcgtcttcgcacaagcttcgtcatcgcgcgcggaagcg 753
Db 718 GATAAAGTTGGAGTAGTTTGGGAAACCTTTTGATCCTTCAACAGGAACAAAGGAAGCCCA 777
QY 754 agtaagaccatttttggggcgtggcgagat-----cccaataatccggtctctctggagcaga 810
Db 778 ACTCAAACTATGTATGTGTGGTGCAGCAGATAAAGACAGAGAAACAATAATTTACGTTACTAAT 837
QY 811 qacggcgcgacgctgacgagcggtgcgcgggtgcgcgacggcgttcacccgcacaaag 870
Db 838 GATGGAGAAAGACATGGAGTGCAGTTTAAAGGACAGCAAAAGGATATCTTCCACATCAT 897
QY 871 ggcgtctttgaccgcgttcaaccgcgtctctatatattgccaccagcaatacgggtgggtccg 930
Db 898 GGTATACTTG--CTTCCGATGATGCTTTATATATCATATATAGCAATACTTGGCGACCA 954
QY 931 tatgagggagctccgcgcgctctggaaattctcgtgacctccgggacatggagcgcga 990
Db 955 TATGACGGCTCAGATGGACAAGTTTGGAACTATAATACCAAAACAGGAGAATGGACAAAC 1014
QY 991 atcagcccggtacctgcagcagcagcgcccaacgactacttgggttacagcgccctcaat 1050
Db 1015 ATAACTCCACCTCGAGTGTGACAC-----TAAATCTGGTTTGGAGGAATATCA 1065
QY 1051 atcagccgcagcagcccaacagataatgtgtgcacccagatatactgtgtgcccagac 1110
Db 1066 GTAGATCCACAAATCCAAATATATGTTGTAGTTGCACACTTTAAATAGATGGTGGCCGGAT 1125
QY 1111 accataatctttggagcaccgcgcggtgcgacgtggagcgcggaactctgggattggacg 1170
Db 1126 GAAGAAATATACAGAAGTACAGATGCAGGTAAAGACCTGGAAACCTATATGGGATTGGAAT 1185
QY 1171 agttatcccaatcgaagcttgcgatatgtgtctgaacttgcgcgagccttggctgacc 1230
Db 1186 GGATATCCAAACCGTACTTTTGGGATAAATTAAGTTACTTACTCAGCACAACTTGGCTTGAT 1245
QY 1231 ttcggcgtaacgcgcg---aatctccgtaccagctccgaagctcggctggatggatgaa 1287
Db 1246 TGGGGTAAAGCCGGGTAAACACCACAGATCCCAATTAGTTAAATTAGGCTGGATGGGA 1305
QY 1288 qcgatggcaatcgatccgttcaactctcgtatcgatcgtctacggaacagcgcgacgttg 1347
Db 1306 GACTTAGAAATTTGATCCATTTACTCGGATAGAAATGTTCTACCGTACAGGTGCAACTCTT 1365

RESULT 5

ABI292929

LOCUS

ABI292929

2579 bp

mRNA

linear

PLN 08-AUG-2000

QY 1348 tacgcaacaaatgatctcacaagtgaggactcccgcgccgagatttcataatcgcgcgcgatg 1407
Db 1366 TATGGAACTGATGATTGTCAATATGGGATAAAGGTAAAGATGTAGACATCTCTGTAAAG 1425
QY 1408 gtcaaaaggattgagagacgcgcgtaaacgatctcatcagcccgctctgcgcgcgcgcg 1467
Db 1426 GCAAAATGGAATTAAGAATGTGAGTAATATGATGTTGTAGTACCAACAAAAGGCGCAAA 1485
QY 1468 ctcatcagcgctcgcgagacccgcgcgcttcaccacccgcgcgcgcgcgcgcgcgcgcgc 1527
Db 1486 CTTTTAAGCGCTGTGGAGATGACTCGGATTTTATCATGATGACATAAATAAGTTTCCA 1545
QY 1528 tcgacgatcttcacgctcaccgcgtgttcacgacccggcaccagcgcgcgcgcgcgcgcgcgc 1587
Db 1546 AGTAAATGATGACACACCTAATCTTAGTCTACACAAAGCATGATTATGACAGAAAGT 1605
QY 1588 aatccgcgcgcatcgttcgcgcgtggaagtctcgatccatcgcgcgcgcgcgcgcgcgcgcgc 1647
Db 1606 GTTCTCAACTTTGTTCTGTTGGAAATGTTGATACTAGTAAATCAACAGATATAA 1665
QY 1648 cacgtcgcgcttcgcagacagc 1704
Db 1666 GATTGTGGAATTTCTACGATGGTGGTAAGAACTGGTCTCAGCTGGAAGCAATATATCA 1725
QY 1705 ggggtgacgagcgccgc 1764
Db 1726 GGAAGTTTAAAGCAGGAAGCTAGCAGCTGGTCTGATGCTAAAACATATAGTATGTCT 1785
QY 1765 ccgc 1824
Db 1786 CCAGAGGAAGTGCTAATGAGCATATAGCACAGATA--ATGGTAATAATGACACCG 1842
QY 1825 tcgcaagggtttccgc 1884
Db 1843 TGTCTGCTCTCCACAGGAGCAAGGTTAGATCAGACGCTGTAATTCAAAAGAAATTC 1902
QY 1885 tatgcctatcccaatgaaccttctcgaagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1944
Db 1903 TATGGTTTCTTAATGGTAATTTCTATATTAGTACTGATGAGTGCAGGTGCAACATTTACTCAA 1962
QY 1945 gtgc 2004
Db 1963 AGTTCACAAACAGGCTCTCTCACAAGGCAAAAGGAATCTTTAAGACTGTTATAGGTCTAT 2022
QY 2005 aaagaaggcgatctgtgctcgtgcatcgc 2064
Db 2023 GAAGGTGATATTTGGATTGCTGGAGGAAAAGACGGATTATGGCATTCACATGATCTCTGGT 2082
QY 2065 agcagttggtctgcaatcaccgc 2124
Db 2083 GCAACATTCACAAAGTATCTGGTTGTGATGCTTCTGTATACAGTAGTCTTGGAAAAATCC 2142
QY 2125 gc 2184
Db 2143 AAACAGATGATGGATATCCAGCTATATACATGGATGCTACAAATTTATGGAACAGCGGGA 2202
QY 2185 gcgtaccgcctccgacgactgtgggacgacccgtgggtactgatcaatgatgaccagcaca 2244
Db 2203 ATATTCGTTTCAGATGATGAAGGTGCAACTTGGTAAGAATAAATGATGATGCTCACCACAA 2262
QY 2245 tacggaattggggaagaagaatcaccgc 2304
Db 2263 TATGGTTCCCGACACTACTGCATAACTTGGAGATCCTTAACAAATATGGACAGATTTTTTTGTT 2322
QY 2305 ggcacaacgc 2350
Db 2323 GGAACTAATGGAAGAGGATTTGTTTATGGAGACATAGATGTTCTC 2368

DEFINITION Agaricus bisporus mRNA for CEL6 protein.

ACCESSION AJ292929

VERSION AJ292929.1 GI:9755896

KEYWORDS cel6 gene.

SOURCE Agaricus bisporus.

ORGANISM

Agaricus bisporus
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Agaricaceae; Agaricus.

REFERENCE 1 (bases 1 to 2579)

Morales-Almora, P. and Thurston, C.F.

Molecular analysis of the cellulolytic genes in Agaricus bisporus

Unpublished

REFERENCE 2 (bases 1 to 2579)

Morales-Almora, P.

Direct Submission

Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's

College London, 150 Stamford Street, London, SE1 8WA, UNITED

KINGDOM

FEATURES

source

1..2579

/organism="Agaricus bisporus"

/strain="D649"

/db_xref="taxon:5341"

/clone="PMRT152"

/dev_stage="mycelium"

/country="United Kingdom"

28..96

/gene="cel6"

28..2448

/gene="cel6"

/function="putative cellulase"

/codon_start=1

/product="CEL6 protein"

/protein_id="CAC02964.1"

/db_xref="GI:9755897"

/translation="MFSRVKFWITLTLGTFSLVYQAAQSQSYRQWVKI GGGGFPVG
IVNPEKGLAYARTDIGGAKLNADDTWPLDFAADSRWNYGVYDALATDPVQPNR
LYLATLYNSWDPNSGHI LSSDYKTFDAAPLPFKLGNMGRGMGRVLYVDPNLN
SILYFGARSGNTGNSGRMSKVSFTDTSFVDPDSPTGLNSDKIGIAWVTID
KASGSGATPRFVGVANKGNSIVISTNGSSWSAVAGQTSFLPHKGVLSPSERA
LXITYSDGGPDPYDTSVAVKYSIDTGWNTDITPVSGDLFTFGFGLDLDOKNCTVM
VAALNSWDPGQIFRSTNGSGASWTPLDWNAVPTLNKYYSYNAALPWLGNLIVTP
GNQIQMESLSIDPFDNSHMLYGTETIYGSRLDKLWDSAHNVTIKSLADGVEETS
VQALISPPGSLVSAIGDVGFFVHQSLTRAPSAQFNPKWSTSDIDFAGNNPSNV
RIGTSGTSQKVAISDGYVTWQHFQAPDNVGGKVAISADADILIRWTNGVMV
SRNOATFNIFGASGTFYVSTDNGKTFSAHSGLSGATSVEDITHPSVSGDIWASTD
KGLFHTSDGSATFSALGITOAWGVALGAPRSTGGYPVAFPAANYGEVAYFRSDDRG
VNWVKINDRAHGFSAASANCMAADPRYGRVYIGTNGRGIFYGDVAGSAPPPTSSAYS
TTSVSTITNRTSITRTTSTTTSSGNGSPVSGFGCGGCGGTGTTCTVTFGQCAQNP
FYSQCIPI"

28..2448

/gene="cel6"

97..2445

/gene="cel6"

/product="CEL6"

BASE COUNT 615 a 662 c 631 g 671 t

ORIGIN

mat_peptide

gene

Query Match 11.4%; Score 327.6; DB 8; Length 2579;

Best Local Similarity 52.6%; Pred. No. 8.3e-28;

Matches 845; Conservative 0; Mismatches 734; Indels 27; Gaps 5;

QY 177 cggggcgcgccgttttcgacgggacgctgttcaatgaagggtgcaccgggaattctgta 236

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 135 CGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 194

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 237 cgtcgacggacatcgggggggtatgatcgatgggatgcccaacggcggtggatccc 296

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 195 TGCTCGCACTGACATCGGAGGAGCCTACAGCTGAATGCTG---ATGACTATTGGAGGCC 251

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 297 tcttctggattgggtggatggacaattgggggtacacggcgcttcagcagcgcg 356

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1320 GGACAGTGCACACACGTCACCAATCAAGTCTCGCAGATGGTGTGCGAGAAACCTCCGT 1379

Db 252 TCACATTGATTTCCGCTGATGATCTCCGTTGGAATATTATTTGGGGCGTCGATGCTCTTGCAAC 311

QY 357 agaccgatcaataactactaacaaggatattgggcgcgcgtcggaatgtacacccaacagctggga 416

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 312 TGATCCCGGTACAACCCCAACCGCTCTTTACCTAGCGACTGGGTTGTATACCAACTCGTGGGA 371

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 417 cccaacgacggagcgatcttcgcctcgtctgacagggcgcaacgctggcgaataaacgcc 476

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 372 CCCAAGTAACGGGCATATTCTCATATCAAGTATTATGTTAAACATTCGATGCCGGGCC 431

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 477 cctgcgcttcaagcttgccgcaacatgcgcggcgctgggaatggcgagcggttcggt 536

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 432 TTTGCCATTCAAGCTCGGTGTAATAATGCTGGCGTGGAAATGGAGAGCGTCTCGCT 491

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 537 ggtatcaaaaatgacaaatcttgatatttcggcgccccgcggagcggaagggctctggag 596

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 492 CGATCCCAATCTGAACACGATCCCTTTATTTGGCGCTCGCAGTGGTAACGGTCTCTGGAA 551

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 597 aagcacagattccggcgcgacaccttcgcagatgacgaacttttcgggacgtaggacgcta 656

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 552 GTCAACAAATTCGGTCCGAGCTGGAGCAAGGTAACAAGCTTCACTGACACTGGGTCTTT 611

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 657 catgcaaatcccaactgacacgacccgctatcagagcgatattcaagcgctcgtctgggt 716

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 612 CGTTCGATCCATCAGATAGCACCGGCTGAAATTTCTGATAAGATTGGGATCGCATGGGT 671

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 717 cgttttcgacaagcttctcgtcgtcgcgcgcaagcgagtaagaccatttttgggct 776

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 672 TACATCGATAAGCGTCAAGCTCCTCTGTTCCGCTACTCCCGGTATATTTGTCGGCGT 731

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 777 ggcg---gatcccaataatccggtcttcttgagcagagcgggcgcgcgctggcgagcc 833

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 732 TCGCAACAAGGCTCAACACAGCATTTATATCTACGAATGGCGGTTCTCTTTGAGAGTGC 791

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 834 ggtgcggggtgcgcgcacccgcttcacccgcaacagggcgcttttgaccccgctcaacca 893

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 792 CGTCGCAAGGTCAACCAACTACTCTCTGCCACATAAAGGGGTTCTATATCCCTCTCTGAGAG 851

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 894 cglgctctatatgccaaccagcaatcggtggtccgtatgagggagctcccgcgacgt 953

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 852 GGCACTTTATATCATTACTTCCGATGGTGGCGGTCCTTACGATGGCAACAGCGGTGAGT 911

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 954 ctggaattctcgtgacctccgggacatggacgcggaatcagcccggtcacttcgcagga 1013

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 912 TTACAAGTACAGCATTTGACACTCGGCATGGACTGACATTAACACCCCTCAGCGGAAGCGA 971

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1014 cagggcaacgactactttgttacagcgccactcactatcgacccgcagcaaccgcaac 1073

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 972 T-----CTCTCTTTTGGTTTCGGAGGTCTAGCCCTTAGATACACAGAAAACGGGAC 1022

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1074 gataatggtggcaaccagatatcgtggtggcgcgacacacataatcttcgggagcaccga 1133

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1023 TGTCATGGTTGCAGCCCTCAACTCGTGGTGGCTGATGGTTCAGATATTCGCTCGACCAA 1082

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1134 cggcggtgcgacgtggacgcggtatctgggattggagcaggttaccgaatcgaagcttgcg 1193

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1083 CGGGGGTGCATCTTGGACTCCGCTGTGGGACTGGGAGTCTACCCGACTCTCAACAAATA 1142

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1194 atatgcttgacatttcggcgagccttggcgaaccttcggcgtacagcgcgaatcctcc 1253

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1143 CTACAGTTAAGCTGCGCTTGGCGCGGTGGATAGGACCAACATCCTTTGATGTACCCC 1202

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1254 cgtaccagtcggaagctcggtggatggatgaagcgatggcaatcgatccgttcaactc 1313

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1203 CGGA---AATCTACAAATCGGTGGATGATGGAATCACTATCCATTGATCTTTCGACTC 1259

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1314 tgatcgatgctctacggaacagcgcgacgtttgtacgcaacaaatgatctcacgaagt 1373

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1260 GAATCATTTGCTCTACGGCACTGGAGAAACTATCTATGGAAGTCTGACCTCTCTCAATG 1319

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1374 ggaactcggcgcgacattcattcgcgcgattgggtcaaaaggattggagagcgcggt 1433

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1320 GGACAGTGCACACACGTCACCAATCAAGTCTCGCAGATGGTGTGCGAGAAACCTCCGT 1379


```

Db 2826 ACAGTGGACGGCGACAAAGCGG---CAGGTGTTCAACTGTGACTGGCGTGGTCCGATGT 2882
QY 2803 ggaatatccgccttcgctcggtgaacccggcgagccgagcgagcctac 2862
Db 2883 TCGAACCTCGCGCGAGCTCTGGTGAAGTTGACGACGGCGCGGCGGCGGACTACTAC 2942
QY 2863 ctgcag 2868
Db 2943 CTTGAG 2948

RESULT 7
LOCUS NPORPT2
DEFINITION Multicapsid nuclear polyhedrosis virus of Orgyia linear VRL 03-FEB-1999
pseudotsugata (OpMNPV), GC-rich repeated DNA from the SstI I
fragment.
ACCESSION D13755 D00297
VERSION D13755.1 GI:222216
KEYWORDS repetitive GC-rich region.
SOURCE Multicapsid nuclear polyhedrosis virus of Orgyia pseudotsugata DNA,
clone cosmid 58.
ORGANISM Orgyia pseudotsugata single capsid nuclear polyhedrosis virus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 749)
AUTHORS Chen,D.D., Nesson,M.H., Rohrmann,G.F. and Beaudreau,G.S.
TITLE The genome of the multicapsid baculovirus of Orgyia pseudotsugata:
restriction map and analysis of two sets of GC-rich repeated
sequences
JOURNAL J. Gen. Virol. 69 (Pt 6), 1375-1381 (1988)
MEDLINE 88259472
COMMENT Analysis of the region from the SstI I fragment that was
crosshybridized with the repeated DNA downstream of the polyhedrin
gene revealed a sequence made up predominantly of repeats of the
sequence GCGC or the closely related variant GGTG, and a
dinucleotide which results in amplification of groups of six
nucleotides.
FEATURES
source 1. .749
/organism="Orgyia pseudotsugata single capsid nuclear
polyhedrosis virus"
/db_xref="taxon:10450"
BASE COUNT 111 a 163 c 289 g 186 t
ORIGIN 5 up upstream of PstI site.

Query Match 3.6%; Score 103.6; DB 14; Length 749;
Best Local Similarity 62.2%; Pred. No. 0.0066;
Matches 163; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 2348 cgcgctcgcgagtcgcgttcctcggtgagtcgcttcgcttcgagcctgagcccg 2407
Db 418 CGCATCGCGAGCCATCACCACCGCGCGCCATCGCGGCGCCATCACCACCGCGCA 359
QY 2408 gccgagccgcgagcagtcgcctcgcgttcgcttcgagtcgagtcctcctcgt 2467
Db 358 CGCCATCGCGCGCCATCACCACCGCGCGCCATCACCACCGCGCCATCGCGCA 299
QY 2468 cgcgctcgcgtcgcgttcaccatcgccgagtcgcttcggttcgcttcgcttcgcttc 2527
Db 298 CGCCATCACCACCGCCATCACCACCGCGCGCCATCGCGGCGCCATCACCACCGCA 239
QY 2528 cgcgagccgcgttcgttcacagcgcgttcgttcacgcttcgcttcgcttcgcttcgcttc 2587
Db 238 CGCCATCGCGCGCCATCACCACCGCGCGCCATCGCGGCGCCATCGCGCGCCGCGCCG 179
QY 2588 cgcgctcgtcgtcgcggtgctc 2609
Db 178 CGCCATCGCGCGCGCGCGCGCC 157
```

```

RESULT 8
LOCUS OPU75930/c
DEFINITION Orgyia pseudotsugata nuclear polyhedrosis virus complete genome.
ACCESSION U75930
VERSION U75930.2 GI:11024985
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and
Rohrmann,G.F.
TITLE The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome
JOURNAL Virology 229 (2), 381-399 (1997)
MEDLINE 9721300
REFERENCE 2 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1996) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
REFERENCE 3 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1998) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter
COMMENT On Oct 26, 2000 this sequence version replaced gi:2934903.
similar to Autographa californica nuclear polyhedrosis virus
(ACMNPV) complete genome: GenBank Accession Number L22858.
FEATURES
source 1. .131995
/organism="Orgyia pseudotsugata single capsid nuclear
polyhedrosis virus"
/db_xref="taxon:10450"
complement(123..947)
/note="ORF1; pk-1, ACMNPV ORF10 homolog"
/codon_start=1
/product="protein kinase 1"
/protein_id="AAC59000.1"
/db_xref="GI:1911247"
/translation="MDAALQSLSRFVADCAVLAPKVVNGRFGKMDVLHHRPTTSKLF
RKTIAAHGFSADEINVDHLSMDHPFVDMYFCYSSTAWAIYMDYVPCDLPFETLQ
GALDNALVYVIRQLCDALNDLHNATGYIHNDVKLENVLYFGARDRVYLCGYCKLRE
HSPVHDGTLVEYSPKIRRHNYARSGFDWYAGVAYKLLTGGRHPPERSADSVLDLAS
MRRQYNDPAAALKVNRNLMARDFVCLTRFNECRSTDYKQIAKHSFLASRDYI"
946..2370
/note="ORF2; ACMNPV ORF9 homolog"
/codon_start=1
/product="1629-capsid"
/protein_id="AAC59001.1"
/db_xref="GI:2934904"
/translation="MMERQYQSVRSYRLINQHNATAAGPFLQRVAGPEAHSVGRNVGD
RAVRLNRQAVLDLLKEDIATYMQADQPEASRHFATLNRRLLLIGVQDPDAR
RNLNSVLARIEALLRVVDVAENVLSGDFYEYSKYISYQQTQAFQTASASQOTQ
TSLRPQTSLPQOTASLPQDPQPEMVSPSPVHTTTPAILPQTPQPDATDTSRP
SDEFYVPGKEKPAVDPKPPVPPKPEHLKSRPSVATNAAGATPAPPPPPPSAD
VTSMPPPPPPSADVTSMPPPPPVVDLATSMPPPPPPPPPPPPVVDLATSMPPPPINNA
INLLIDAMVAETKNAGNRSALLDQIKQGLTKRKTQPDGAPDPPSTLLSEIRQ
KRRPAVEHTDGNSTGNSDWRDD"
complement(2533..3270)
/note="ORF3; ph, ACMNPV ORF8 homolog"
/codon_start=1
/product="polyhedrin"
```


misc_feature 2033..2884
/gene="mana"
/note="Region: beta-1,4-mannanase catalytic domain"
/evidence="experimental"
2885..2959
/gene="mana"
/note="Region: proline-threonine linker"
/evidence="not_experimental"
2960..3409
/gene="mana"
/note="Region: cellulose-binding domain type IIIB"
/evidence="not_experimental"
3894..4400
/note="orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAP22275.1"
/db_xref="GI:6651328"
/translation="MGASGVFSETAGNYGLKARCKRRCVGNLLGHLRFVVDDEF
RFAGARKRLGRFSRNRNLGEAQNFESEVTFFVDDRRIRRRNGLGFVRFAQSI
GPKRLIEIAEQARAKQASDPQLKNVILHFDIHPYGVVERMDPFHFKIEKOKS
RSLITRY"
BASE COUNT 1040 a 1276 c 1439 g 812 t
ORIGIN

. Query Match 3.4%; Score 96.4; DB 1; Length 4567;
Best Local Similarity 57.1%; Pred. No. 0.027;
Matches 216; Conservative 0; Mismatches 156; Indels 6; Gaps 2;
QY 2494 ccgagtcgtctcgttcggtccaccatcggtcgccgagccggtcttcgtcaccagc 2553
Db 2837 CCGACGCGTGGGGGACATGGTCCCGAGCAACGCGATCGGGACCTCGACTTCGCCGACA 2896
QY 2554 ccgtcttcgtcaccgtcttcgtcgccgagcccaacgcgcgtcgctcgccggtgctcggt 2613
Db 2897 CCGACACCTACACGCGCGCGACGCCAACCGCGACCGACCGCGCTCGCGCGCG 2956
QY 2614 gggg---tgaagtgagatagaataatgattcgcgccggtgataatcgatacaag 2670
Db 2957 GGCAACTTGTGCTACAGTACCGCGCGCGGACACGACGACGACGACGACGACGCTGAAG 3016
QY 2671 ccgggttcgagtggtggaataccggttcgtcggtgattggtcgagcggtgacggtg 2730
Db 3017 CCGCATTTTCGATGTGACACCGCGGACGCTGAGCGTGTGCGGACGCTGACGATT 3076
QY 2731 cggtaactggttcacccggggtggtggtcgctcgacactggtgtacaaactgtgactggcg 2790
Db 3077 CGGTACTGGTACACGGGTGGACGGGACAAACCG--CAGGTGTCAACTCGGACTGGGCG 3133
QY 2791 gcgacggtggtggaataatccgcgctcgttcggtcggtgaacccggcgagccgacg 2850
Db 3134 CAGGTGGGTTGCTCGAACTTCGCGGGGACGCTGAGCGTGTGCGGACGCTGCGGCGG 3193
QY 2851 gcgacacccactacgtcgag 2868
Db 3194 CGGACTACTACATCGAG 3211

RESULT 11
AL645600/C

LOCUS AL645600 183665 bp DNA linear HTG 01-FEB-2002
DEFINITION Mus musculus chromosome 11 clone RP23-396N4, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL645600
VERSION AL645600.8 GI:18491381
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (sites)

AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 4, 2002 this sequence version replaced gi:18476868.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm396N4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 183201 bases at least Q40
Consensus quality: 183287 bases at least Q30
Consensus quality: 183384 bases at least Q20
Insert size: 183465; sum-of-contigs
Insert size: 176789; 3.5% error; agarose-fp
Quality coverage: 13.90x in Q20 bases; sum-of-contigs Quality
coverage: 15.10x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..183665
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-396N4"
/clone_lib="RPC1-23"
1..3150
/note="assembly_fragment:05886"
misc_feature 351..60564
/note="assembly_fragment:00037
fragment_chain:1"
60665..183665
misc_feature /note="assembly_fragment:06038
fragment_chain:1
clone_end:SP6
vector_side:right"
BASE COUNT 51045 a 35803 c 35999 g 60618 t 200 others
ORIGIN

Query Match 3.2%; Score 91.4; DB 2; Length 183665;
Best Local Similarity 56.0%; Pred. No. 0.04;
Matches 173; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 2347 gcgccgtccggatcgccgtctccgtcggtgagtcgctcgcttcgagccgagcctgagccg 2406
Db 127835 GCTCCGGCTCCAGTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127776
QY 2407 agcccgagcccgagcagctcgccatcgccgtcgccgagcagctcgagtccttcctcg 2466
Db 127775 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127716
QY 2467 tcgccgtctcgtcgccgtccaccatcgccgagtcgctcgttcgttcgttcgttcgttcgttcg 2526
Db 127715 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127656
QY 2527 tcgccgagccgttcgttcgacgagcccgcttcgttcgacgcttcgttcgttcgttcgttcgttcg 2586
Db 127655 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127596
QY 2587 acgccgtcgttcgctcgccggtggtcggtgggtgaagtgtagtataagaataatgattcg 2646
Db 127595 GCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCGGCTCCG 127536

Qy 2647 gcgcgggt 2655
 || |||| |
 Db 127535 GCTCCGGCT 127527

RESULT 12

AP003605 40340 bp DNA circular BCT 05-DEC-2001
 LOCUS Nostoc sp. PCC 7120 plasmid pCC7120epsilon DNA, complete genome.
 DEFINITION Nostoc sp. PCC 7120 plasmid pCC7120epsilon DNA, complete genome.
 ACCESSION AP003605
 VERSION AP003605.1 GI:17134931

KEYWORDS
 SOURCE Nostoc sp. PCC 7120 plasmid:pCC7120epsilon DNA.
 ORGANISM Nostoc sp. PCC 7120

REFERENCE 1 (sites) Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

AUTHORS Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S.,
 Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
 Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
 Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
 Yasuda, M. and Tabata, S.

TITLE Complete genomic sequence of the filamentous nitrogen-fixing

JOURNAL cyanobacterium Anabaena sp. strain PCC 7120

MEDLINE DNA Res. 8 (5), 205-213 (2001)

PUBMED 21595285

11759840

REFERENCE 2 (bases 1 to 40340)

AUTHORS Kaneko, T.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
 Institute, The First Laboratory for Plant Gene Research; Yana
 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail: kaneko@kazusa.or.jp,
 URL: http://www.kazusa.or.jp/cyanobase/.

FEATURES

source Location/Qualifiers

1. 40340

/organism="Nostoc sp. PCC 7120"

/plasmid="pCC7120epsilon"

/db_xref="taxon:103690"

/note="synonym: Anabaena sp. PCC7120"

1429. 3222

/gene="alr9001"

/note="ORF ID: alr9001"

similar to mobilization protein"

/codon_start=1

/transl_table=11

/protein_id="BAB77487.1"

/db_xref="GI:17134932"

/translation="MVAAIILRVKLSFGNGSEKHTARLQDTNADTTKKNIRLI

GMEDNSLEDVYKTKANTKKPKDAVLCSEMFLSASPEYERDDPSRAGWNHER

MVLFACASRWLNNYGDCKVRAELHDEATPHIAYIPIINDTKQLSKHEMFGNG

RVGSILSKLQDSALAPLGIERGKGSKATIKVKEYIQAVNSEPLTAVSNKLL

APQLESATNYARIQNDQDFHAINQLADRAPMQERLEAQRARASEKERQLEKE

VRSELTKQLRDLLELDVAVELGLDYERERWGHGHIINIDGSKFYDFSPQKQGG

GADLVHMVQCNRQAVVWLHERFGEAGVERAAIAHVKNRAADIQAEPQPTPPV

ENKANMPAVERYLTQQRGIPSDCLQMLGLLYADQQNAVFMVMDLEGQSGAFLR

GTFCENNSPKYKSKTKRSDSMFYFCLGQATDKVAHVLLCSPPIEAVSRAMLEYVR

GNNVPRTLYMVDNINSLPRLQVNPNIYTFGDKQSTNAQAQVLELLPQSQQVL

SKAADWNQQLLEGWLRQYQQQDDLSI"

complement(3399..3833)

/gene="alr9002"

complement(3399..3833)

/note="ORF ID: alr9002"

unknown protein"

/codon_start=1

/transl_table=11

/protein_id="BAB77488.1"

/db_xref="GI:17134933"

gene

CDS

/translation="MRKLVPSTDERFYRELKTAQAQEEVEVITDVERYSDLPOLKTI
 IFELHNKSGMVNVATGAFIPYSFAATTINYSALYIFGGAGAGATFVGVVGGPVGAA
 VGGGIGAIIVGTAVATLNGHKHRIEINAGNKLRFKISPSIK"
 complement(4726..8019)
 /gene="all9003"
 complement(4726..8019)
 /gene="all9003"
 /note="ORF_ID: all9003"
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77489.1"
 /db_xref="GI:17134934"
 /translation="MVAQSSSTAIAGSDFIIRNFIDQLEPTRVKNKYICPVCGGHN
 LSIINPNNGKSYCNEQCHLRDIRAIPWTQVLEERKLGSITSPKPLPIKAKKPAIVPK
 VLDVDPQLRIICLLGEITTPQVTPDFPKSVAIRLSDSGATSQLKETKETEYDYG
 NGRKAHRFSCPCAAAPKGRVKTFSVRIDPITNKVAKKGEFWPAXRQSEAIATIKAT
 DGIPLVLAHEGKCVASRLNLSITWVGSSSDRDLISLTOIQHSTGKDFLLAYCV
 DNDSQWNKQRIKEICQQAQVSVFVAILDLKIQPNLQDKGDIADIESGTMGDELVEL
 LKQIBEITQKQQAQDTLDVADLDEEDPSTFLQKGFNTLYGETKWTACDGLVY
 WHGKYRSPDAVERRRITNYCNSYPIKKNQSDYVITYPYAKPNKVKELEWKLRY
 EIDPQLNPNVGNCTNGVLAIDWSGPIPRPVEHPDKHFTFYELPKVYDKPADSSH
 CDRLLCLDSPOQOILLNLGASLDLAELAEVRKRGPKVLLACGLGANGKDSIQVVS
 TIYHOGMTSCSLADFAVDEGRKFALALVNSRVNASENPQTTRLDKIQSLKLEAT
 GNVLSERKGDHIEFNSNAIGLNFNDTPSWTGTQAILDRIALIFRFTKTFNDPP
 NNHDELLADPRFAYDVFQECVAPAFNLNKMIAAGLQALIAEGIDYISCTQALBEIQAE
 NSHLFQCKDTGLGYKANGIVTAFDIQRLNNYIDNGTLSPFESSNGKVAIKWQS
 KPSDKTVAINOALPRIKALFPKTKLVTPHPSGKRQALQALQIGISFDNVPMSITGPI
 ATSENSTPNSTPNHQSLINQDFTHTTSTDEGLDEKKNNGFQVLDQDQPKSEIINN
 VDRADQWHGNLESKLIDINPNDVFENEIEGDKLWVEVCEVPASTTADSTGVDIAPVG
 ELTAMPELSLTTEVESQLKSDVAEVLQGEAEASHQTVNQVDEQPKSAFQSPCS
 TSGSTPELSFEQKLMGNVALIRECIADESGLIAELIEETWTSFKSAVWKELTPEERQ
 AVKQLKALE"
 complement(8400..8570)
 /gene="parg"
 complement(8400..8570)
 /gene="parg"
 /note="ORF_ID: asl9004"
 parg gene product"
 /codon_start=1
 /transl_table=11
 /product="plasmid partition protein"
 /protein_id="BAB77490.1"
 /db_xref="GI:17134935"
 /translation="MTDSNKKQVFRVDEKRDREKFKITCIQLKTTTMDVYLKGLLDKW
 LEENDPESTKTK"
 8828..9880
 /gene="alr9005"
 8828..9880
 /gene="alr9005"
 /note="ORF_ID: alr9005"
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77491.1"
 /db_xref="GI:17134936"
 /translation="MTDLAEMPENALPRHQGTULTIAGYDLGNSGVKFTVSDRKIRFP
 SYLXENYRPTLPTGYEYLEGDAITKLDYKQWLSGYAAVDANPKNLHRTVDDATA
 KYTQSLKHLAALSNNPYKPVINLIICASHERGDLLEQTLIDAIGKHLVKGFGKPI
 TVNIVHLVYDEGHAAIAANAHTLDTSKQNVTDIGNRTVTLTLGQKHLANRKT
 DNGVERLIMISVNPFTKRLYGEIAPHILRGLESSEKPFYHQQSFEDYVRQEL
 MPWQKSLAPVFKFIHPWKINADACLIIIGGSGOLPSVDEALKAKGEFVIAENPLWANA
 E"
 9921..10088
 /gene="asr9006"
 9921..10088
 /gene="asr9006"
 /note="ORF_ID: asr9006"
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB77492.1"
 /db_xref="GI:17134937"

gene

CDS

gene

CDS

JOURNAL

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625700.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCSW
Center clone name: CH230-8G23
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList

Consensus quality: 148905 bases at least Q40
Consensus quality: 156308 bases at least Q30
Consensus quality: 160895 bases at least Q20
Estimated insert size: 150335; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1
10936: contig of 10936 bp in length
10937 11036: gap of unknown length
17011: contig of 5975 bp in length
17012 17111: gap of unknown length
17112 17111: gap of unknown length
21819: contig of 4708 bp in length
21820 21919: gap of unknown length
21920 27267: contig of 5348 bp in length
27268 27367: gap of unknown length
27368 32998: contig of 5631 bp in length
32999 33098: gap of unknown length
33099 38238: contig of 5140 bp in length
38239 38339: gap of unknown length
38339 42861: contig of 4523 bp in length
42862 42961: gap of unknown length
42962 48627: contig of 5666 bp in length
48628 48727: gap of unknown length
48728 52551: contig of 3824 bp in length
52552 52651: gap of unknown length
52652 56759: contig of 4108 bp in length
56760 56859: gap of unknown length
56860 60122: contig of 3263 bp in length
60123 60222: gap of unknown length
60223 64019: contig of 3797 bp in length
64020 64119: gap of unknown length
64120 67830: contig of 3711 bp in length
67831 67930: gap of unknown length
67931 70511: contig of 2581 bp in length
70512 70611: gap of unknown length
70612 74478: contig of 3867 bp in length
74479 74578: gap of unknown length
74579 77220: contig of 2642 bp in length
77221 77320: gap of unknown length
77321 80928: contig of 3608 bp in length
80929 81028: gap of unknown length
81029 83523: contig of 2495 bp in length
83524 83623: gap of unknown length
83624 86504: contig of 2781 bp in length
86405 86504: gap of unknown length
86505 89767: contig of 3263 bp in length
89768 92456: contig of 2589 bp in length

92457 92556: gap of unknown length
92557 94688: contig of 2132 bp in length
94689 94789: gap of unknown length
94790 97737: contig of 2948 bp in length
97738 97836: gap of unknown length
97837 100283: contig of 2447 bp in length
100284 100383: gap of unknown length
100384 103335: contig of 2952 bp in length
103336 103435: gap of unknown length
103436 106003: contig of 2568 bp in length
106004 106103: gap of unknown length
106104 108849: contig of 2746 bp in length
108850 108949: gap of unknown length
108950 112359: contig of 3410 bp in length
112360 112459: gap of unknown length
112460 116805: contig of 4345 bp in length
116806 116905: gap of unknown length
116906 118553: contig of 1948 bp in length
118554 118953: gap of unknown length
118954 121180: contig of 2227 bp in length
121181 121280: gap of unknown length
121281 123672: contig of 2392 bp in length
123673 123772: gap of unknown length
123773 125124: contig of 1352 bp in length
125125 125225: gap of unknown length
125226 126338: contig of 1114 bp in length
126339 126438: gap of unknown length
126439 127933: contig of 1495 bp in length
127934 128033: gap of unknown length
128034 129916: contig of 1883 bp in length
129917 130016: gap of unknown length
130017 132151: contig of 2135 bp in length
132152 132251: gap of unknown length
132252 133407: contig of 1156 bp in length
133408 133507: gap of unknown length
133508 134852: contig of 1345 bp in length
134853 134952: gap of unknown length
134953 137209: contig of 2257 bp in length
137210 137309: gap of unknown length
137310 138426: contig of 1117 bp in length
138427 138526: gap of unknown length
138527 139723: contig of 1197 bp in length
139724 139823: gap of unknown length
139824 141817: contig of 1994 bp in length
141818 141917: gap of unknown length
141918 143387: contig of 1470 bp in length
143388 143487: gap of unknown length
143488 145552: contig of 2065 bp in length
145553 145652: gap of unknown length
145653 147296: contig of 1644 bp in length
147297 147396: gap of unknown length
147397 148991: contig of 1595 bp in length
148992 149091: gap of unknown length
149092 150398: contig of 1307 bp in length
150399 150498: gap of unknown length
150499 152456: contig of 1958 bp in length
152457 152556: gap of unknown length
152557 154951: contig of 2395 bp in length
154952 155051: gap of unknown length
155052 156913: contig of 1862 bp in length
156914 157013: gap of unknown length
157014 159194: contig of 2181 bp in length
159195 159294: gap of unknown length
159295 161394: contig of 2100 bp in length
161395 161494: gap of unknown length
161495 163410: contig of 1916 bp in length
163411 163510: gap of unknown length
163511 164613: contig of 1103 bp in length
164614 164713: gap of unknown length
164714 166031: contig of 1318 bp in length
166032 166131: gap of unknown length

Query Match

3.0%; Score 86.4; DB 2; Length 183544;


```

gene      complement(2104.. .2715)
/ gene="lef-2"
CDS      complement(2104.. .2715)
/ gene="lef-2"
/ note="ORF4"
/ codon_start=1
/ product="LEF-2"
/ protein_id="AAK85568.1"
/ db_xref="GI:15213129"
/ translation="MDNVAVNPSANTATIKKNEYTLDPNDPMDVLTLSPTVTPRRG
OLVEISGRMLFALTMTSPTEPKQOFARRSKNICLKACADGSQLHKSLSLGLKPLPC
IAKIMAEISNASVPRGMYRKRFEFCVGLGNVLTCTCKCKSACLIGALLHFFYKMDAKCV
GEVTHLLKSDQVYKPSNCAKMSKMLCPQANMCKGLNPICN"
CDS      complement(2717.. .3070)
/ note="ORF5"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85569.1"
/ db_xref="GI:15213130"
/ translation="MIYRYIYFFVFRESRTKSKKPRISHIKLFLFYATLHYNNRR
INAPVMVAGHDYDREQLKRDINSRHSVHELCKRSTTGFCNRLLESSTDAKPTVV
IKTTAAEQHTCDKL"
CDS      2990.. .3424
/ note="ORF6"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85570.1"
/ db_xref="GI:15213131"
/ translation="MADARFLRFSTRLEYKENVVAQIDHLTNLRLINAKTTIANV
RFRGVDFNAFVACMDIVHTFYASSDDKITLQEPOLYFRVRCFECNEIADVPDDHS
TRYLCAACGCLVDNPLNVTEDGVQRFMEIQINAGGE"
gene      complement(3426.. .3935)
/ gene="ptp-1"
CDS      complement(3426.. .3935)
/ gene="ptp-1"
/ note="ORF7"
/ codon_start=1
/ product="PTP-1"
/ protein_id="AAK85571.1"
/ db_xref="GI:15213132"
/ translation="MFPDRHDYTVCGNVIEGTNFCFVPLREEMFEYVTVNDLRWT
IKHLKQCSGLGAVDLANTFRYDANMRDEGLLYKKIRVPGVLPDENVQVLFEDV
VKHFTARGPGLIGVHCETHGLNRTGYLCVRYMINILINISPDQAIARFETARGHKTERQ
NYIQHLHL"
CDS      4168.. .5145
/ note="ORF8"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85572.1"
/ db_xref="GI:15213133"
/ translation="MSLSKLLVYDYYGNSMCGVNEQYGAHYLKYIVDEHITNSYVG
DYPWMHREIDTARQILSGALCVNDACHMLDLDLDTKRLATWYKCGNTKGVCHEVYVL
KAIDLVPIDKRSGQGVDDVDFDTILLINTKLPEVLHQLYRFRVYQKLEHVTDFVN
PSADRVGWYKQFVLTMYRILINKSVPAEELLTRQKAVVEYIQPNAGDKYCAEYMA
NYYGFCGIGKHQFAEHTTANMHIMFKSMRGOTTWVDERYKSFVTKDFSRCKETYL
DLKSHVDSLYINCTDKOKNAMEDLLCCQSEQDLDDVCYDIYNMLFFK"
CDS      complement(5163.. .5516)
/ note="ORF9; unique to Eppomnpy"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85573.1"
/ db_xref="GI:15213134"
/ translation="MSSNSFMVDVNTHENVETNVKDTLLKPINWNSNDPSQARCEK
GAPPYFAYKCVPLESPURGNQYKIDPTLPALOKITTTDTANKTPVLASVKPNNYQDN
TVAEAYNNYKNNIT"
CDS      complement(5531.. .6490)
/ note="ORF10"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85574.1"
/ db_xref="GI:15213135"
/ translation="MFDMMFGWMSFNEQHQHHPPLANHFNAODYKOYAVDRRAQSDLV
NRSVKFCHPFTFKRYVIDDDNGRCRVDVDFCKGLEINHELMNCKWDSKHYHLNEI

```

Query Match 2.9%; Score 83.6; DB 14; Length 118584;
 Best Local Similarity 57.1%; Pred. No. 0.32;

```

VFKTPPEISPDMSGTVATKTHGLIQLQQLSFVEYKDDVLLAIKTKDGYCCDDVRDNI
KTVLKHITLNVNSDKFINAHUKLENQVCARPEQLEQRLETILERVDPDAPMFGVLPFR
DYNKHQHLAVFNQERNTQIGFARGQEEYFRKRKLEFEEDDMHMKLETVHPNQMAV
QCIKDRFISNGYKIKKSVNRRVIEVDCNINAAKDIVNNVIIS"
complement(6427.. .7182)
/ gene="lef-1"
/ gene="lef-1"
/ complement(6427.. .7182)
/ gene="lef-1"
/ note="ORF11"
/ codon_start=1
/ product="LEF-1"
/ protein_id="AAK85575.1"
/ db_xref="GI:15213136"
/ translation="MATWYTLDRVNLWMNSYANNRSKFAEFTLKQRIWHYDRHFDN
ALQLLNFINSNSVDVHVKPLDDEGREWVIDADFKNYNKTELMKINVGATALLLFF
TEMHVHRIFMFSGNRGFHMLKPADGFKALKSVREHWYRIFEKTRLNMRDIRGSE
AHCQRAVMYQDQILQSSKPDLELHFWDKDVFCNPTKQIRAPESYNHKGTOFS
RCITQOLLEKIQKSCGSGDGGASTSSISNITH"
gene      7254.. .8732
/ gene="egt"
CDS      7254.. .8732
/ gene="egt"
/ note="ORF12"
/ codon_start=1
/ product="EGT"
/ protein_id="AAK85576.1"
/ db_xref="GI:15213137"
/ translation="MHFAVLLTLLAAQAAHANILAVLTPAYSHHLVYKAYVOALANK
CHNVTVIKPQLLYVDSNAHMGCHIEQLEADMSQYKRLKLVSSGRRGVVADETT
TVANNYLMDFMDKQNNVNVQOFLNCRTEFDVVVEAFADYLVFGHLFRPAPVIO
IAPGGLDENEDTAGAVARHPKYYPNIWRFSTGVGTAGALSEWRLYNEPELLARYSDK
LLKIQFGPTPTPIRELRNNVQLLLNLHPYDNNRPVPSVOYLGGLHLTQTQPNL
QKTLERQLNASVNGIIVYVSGSIDTKSIHDEFQLMLVNTFTGLNNRTVLMKVVDAVY
DSIKLPNNVITQNNFNRQAVLHKNTVAFVTQGLQSSDEALHAQVPMVCLPMGDQF
HANKLQEFKVARTLNTVDVSQILAITDVIIVNKKIYQTRMAELRAVIDYDEIAPA
DKAIKTERVIRKFGHDITYPARSLKSPAANLDQSNVYFISFPL"
8914.. .9528
/ note="ORF13"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85577.1"
/ db_xref="GI:15213138"
/ translation="METPKMLPSSFATPKRAALGTFVKTVTVTVTTSVSGRLTRDERNR
IFQVIAQLKRTLSLTKLTQLQKVRNMQRVLRKNNIILADLAQLERRCRSGSK
YFAVICQGLVITISGQFVQRVANCAVGEQIFCERRNDCHADROLIIABALAA
LGSDVITSAANKRFRKILNVKVVNAKFIVQOQLHNGFNPNYPTH"
9497.. .10123
/ note="ORF14"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85578.1"
/ db_xref="GI:15213139"
/ translation="MDLTITLIPISLKNIEPKRSECFKLTSMREDAEFCFLNVKCRSP
FAGKVLISITFNNEHLQATVCSQVDSVIMNPSQOEIIFDGTIKHDEGVTVPFI
VGPLFSLCESVMPHVHATVDAIRLQVTKVFINIYALKGARTLKLKLFSDNNESD
LVNNIVNFTNDKIECEKYVNTKVPAINVYTGKQLTLVLFIFKN"
complement(10162.. .11235)
/ note="ORF15"
/ codon_start=1
/ product="unknown"
/ protein_id="AAK85579.1"
/ db_xref="GI:15213140"
/ translation="MDCVANQVCSGTLPIVYVNTREDLCLRNKITAKIGLRFFKCFEA
VTADKSGFLVLSGAAACHIDEKDRMLKCLDFDYNCSEKWLQLAQMOKHLOACVDV
AFDNLTLQNTANIMQSPILTILKCFQNGAYCFGNDDVNLRLPLHIDTIRTFNNEFDLI
REALQVEMWQTVGDEYNNKMLINKRHAVIENFVFNVRMKHPFGERCCKKIFMF
GNAVHVLVSPHLRLNDQIMCLIKDIFTDLKLEFVERRTLIRALFLKLPQDACDVCV
NNHSDVAVGKHKREFNITSECKKTLDIYGALGCRKLVIYALTTNTFLNPIHVYADYA
NYPHNRQSCQEQKREFMGVNYLL"
11242.. .11562

```


Matches 152; Conservative 0; Mismatches 114; Indels 0; Gaps 0;	
QY 2348	cgccgtccggtatcgccgtctccgtctccgtgagtcggtcgccgagcctgagcccca 2407
Db 64615	CCCCCTCGCCCAACGGCTCCACCATCGCCAAACCCCTTCGCCAAGCCTCCACCATCGCCAA 64556
QY 2408	gcccagagcccgagcagctcgccatcgccgtcgccgtcgccgagctgagtcctatcctcgt 2467
Db 64555	CCCCCTTCAACCAACCCCTTCGCCAAGCCTCCACCATCGCCAAGCCCTTCACCAACGCCGT 64496
QY 2468	cgccgtctcggtcgccgtcaccatcgccgagtcggtctcggtctcggtcaccatcgccgt 2527
Db 64495	CGCCAAACACCGTCGCCAAGCCGTCGCCGTCTCTTTGCCCTTCAACACCTTCGCCAA 64436
QY 2528	cgccgagcccgctcttcgtcaccgagcccgctcttcgtcaccgctcttcgtcggcagcccaa 2587
Db 64435	CGCCGTGCGCCAAACACTATGCGCAACACCTTCGCCAAGCCCTTCTCCAACACTTTGGCCAA 64376
QY 2588	cgccgtcggtcgccggtgtcggtt 2613
Db 64375	CACCTTCGCCGTACCTTCGCCCAAGT 64350

Search completed: July 2, 2002, 11:13:18
Job time: 7304 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:38 ; Search time 307.67 Seconds
(without alignments)
16010.100 Million cell updates/sec

Title: US-09-917-376-2

Perfect score: 2869

Sequence: 1 atgagctgttcgagaacat.....ggcgagacactacctgcagn 2869

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	100.8	3.5	390	13	AAQ21833
C 2	100.8	3.5	390	14	AAQ36859
C 3	100.8	3.5	390	22	AAF76910
C 4	89	3.1	784	21	AAQ47999
C 5	89	3.1	2011	21	AAQ52204
C 6	86.4	3.0	37286	23	AAQ55922
C 7	85.4	3.0	777	21	AAQ44444
C 8	85	3.0	796	21	AAQ45005
C 9	80	2.8	11706	20	AAQ55661

10	80	2.8	11707	24	AAQ26525	Active cellulase p
C 11	79	2.8	4403765	22	AAI99683	Mycobacterium tube
C 12	78.2	2.7	58857	21	AAA58471	Nucleotide sequenc
C 13	76.8	2.7	651	23	AAQ84956	DNA encoding novel
C 14	72.8	2.5	6510	22	AAQ45648	Tumour suppressor
C 15	72.8	2.5	6510	24	ABL32395	Human immune syste
C 16	72.8	2.5	6510	24	AAQ61115	Human gene regulat
C 17	72.6	2.5	38186	20	AAQ32028	Human METHI relat
C 18	72.6	2.5	38186	22	AAQ90085	Myobacterium spec
C 19	71.4	2.5	1164	20	AAQ34250	Myobacterium spec
C 20	71.4	2.5	1165	20	AAQ34249	Myobacterium spec
C 21	71.4	2.5	1166	20	AAQ34248	Myobacterium spec
C 22	71.2	2.5	494	19	AAV64542	M. tuberculosis im
C 23	71.2	2.5	494	19	AAV44433	Myobacterium tube
C 24	71.2	2.5	494	20	AAZ19343	M. tuberculosis an
C 25	71.2	2.5	494	20	AAZ19131	M. tuberculosis re
C 26	70.2	2.4	1596	17	AAQ22316	Nocardiosis sp. p
C 27	70.2	2.4	4403765	22	AAI99683	Myobacterium tube
C 28	70.2	2.4	4411529	22	AAI99682	Myobacterium tube
C 29	70	2.4	390	13	AAQ21833	Randomising oligon
C 30	70	2.4	390	14	AAQ36859	PCR primer for 5'
C 31	70	2.4	390	22	AAF76910	Sequence containin
C 32	68.8	2.4	18596	22	AAQ31109	Thymidylate synth
C 33	68.8	2.4	18596	22	AAQ91215	Human thymidylate
C 34	68.2	2.4	3543	23	ABL16717	Drosophila melanog
C 35	68.2	2.4	13837	23	ABL16716	Drosophila melanog
C 36	67.8	2.4	407	19	AAV64539	M. tuberculosis im
C 37	67.8	2.4	407	19	AAV44430	Myobacterium tube
C 38	67.8	2.4	407	20	AAZ19340	M. tuberculosis an
C 39	67.8	2.4	407	20	AAZ19128	M. tuberculosis re
C 40	66.6	2.3	1766	19	AAV04490	Corn p-hydroxyphen
C 41	66.6	2.3	2293	19	AAV09659	A. cellulolyticus
C 42	66.4	2.3	6415	20	AAQ55662	DNA sequence encod
C 43	66.4	2.3	6416	24	AAQ26526	Active cellulase p
C 44	66.2	2.3	1294	20	AAQ206824	Streptomyces albid
C 45	66.2	2.3	1294	22	AAH74537	Nucleotide sequenc

ALIGNMENTS

RESULT 1	AAQ21833/c
ID	AAQ21833 standard; DNA; 390 BP.
XX	AAQ21833;
AC	AAQ21833;
XX	08-JUN-1992 (first entry)
DT	Randomising oligonucleotide used in SPRT mRNA prepn.
DE	Systematic polypeptide evolution by reverse translation; SPRT;
XX	ligand binding; ss.
KW	Synthetic.
XX	WO9202536-A.
PN	20-FEB-1992.
XX	01-AUG-1991; 91WO-US05463.
PF	02-AUG-1990; 90US-0561968.
XX	(COLS) UNIV OF COLORADO.
PA	Gold L, Tuerk C;
PI	WPI; 1992-080018/10.
XX	New method of systematic polypeptide evolution by reverse
PT	translation - by linking each polypeptide in sample mixt. to
PT	individualised mRNA allowing further synthesis of selected

KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX	Arabidopsis thaliana.		
PN	EP1033405-A2.		
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.	24-JUN-1999;	99US-0140695.
PR	05-MAR-1999; 99US-0123180.	28-JUN-1999;	99US-0140823.
PR	09-MAR-1999; 99US-0123548.	29-JUN-1999;	99US-0140991.
PR	23-MAR-1999; 99US-0125788.	30-JUN-1999;	99US-0141287.
PR	25-MAR-1999; 99US-0126264.	PR 01-JUL-1999;	99US-0141842.
PR	29-MAR-1999; 99US-0126785.	PR 01-JUL-1999;	99US-0142154.
PR	01-APR-1999; 99US-0127462.	PR 02-JUL-1999;	99US-0142055.
PR	06-APR-1999; 99US-0128234.	PR 06-JUL-1999;	99US-0142390.
PR	08-APR-1999; 99US-0128714.	PR 08-JUL-1999;	99US-0142803.
PR	16-APR-1999; 99US-0129845.	PR 09-JUL-1999;	99US-0142920.
PR	19-APR-1999; 99US-0130077.	PR 12-JUL-1999;	99US-0142977.
PR	21-APR-1999; 99US-0130449.	PR 13-JUL-1999;	99US-0143542.
PR	23-APR-1999; 99US-0130510.	PR 14-JUL-1999;	99US-0143624.
PR	28-APR-1999; 99US-0130891.	PR 15-JUL-1999;	99US-0144005.
PR	30-APR-1999; 99US-0132048.	PR 16-JUL-1999;	99US-0144085.
PR	04-MAY-1999; 99US-0132407.	PR 16-JUL-1999;	99US-0144086.
PR	05-MAY-1999; 99US-0132485.	PR 19-JUL-1999;	99US-0144325.
PR	06-MAY-1999; 99US-0132486.	PR 19-JUL-1999;	99US-0144331.
PR	06-MAY-1999; 99US-0132487.	PR 19-JUL-1999;	99US-0144332.
PR	07-MAY-1999; 99US-0132863.	PR 19-JUL-1999;	99US-0144333.
PR	11-MAY-1999; 99US-0134256.	PR 19-JUL-1999;	99US-0144334.
PR	14-MAY-1999; 99US-0134218.	PR 20-JUL-1999;	99US-0144352.
PR	14-MAY-1999; 99US-0134219.	PR 20-JUL-1999;	99US-0144632.
PR	14-MAY-1999; 99US-0134370.	PR 20-JUL-1999;	99US-0144884.
PR	18-MAY-1999; 99US-0134768.	PR 21-JUL-1999;	99US-0144814.
PR	19-MAY-1999; 99US-0134941.	PR 21-JUL-1999;	99US-0145086.
PR	20-MAY-1999; 99US-0135124.	PR 21-JUL-1999;	99US-0145088.
PR	21-MAY-1999; 99US-0135353.	PR 22-JUL-1999;	99US-0145085.
PR	24-MAY-1999; 99US-0135629.	PR 22-JUL-1999;	99US-0145087.
PR	25-MAY-1999; 99US-0136021.	PR 22-JUL-1999;	99US-0145089.
PR	27-MAY-1999; 99US-0136392.	PR 22-JUL-1999;	99US-0145192.
PR	28-MAY-1999; 99US-0136782.	PR 23-JUL-1999;	99US-0145145.
PR	01-JUN-1999; 99US-0137222.	PR 23-JUL-1999;	99US-0145218.
PR	03-JUN-1999; 99US-0137528.	PR 23-JUL-1999;	99US-0145224.
PR	07-JUN-1999; 99US-0137724.	PR 26-JUL-1999;	99US-0145276.
PR	08-JUN-1999; 99US-0138094.	PR 27-JUL-1999;	99US-0145913.
PR	10-JUN-1999; 99US-0138540.	PR 27-JUL-1999;	99US-0145918.
PR	10-JUN-1999; 99US-0138847.	PR 27-JUL-1999;	99US-0145919.
PR	14-JUN-1999; 99US-0139119.	PR 28-JUL-1999;	99US-0145951.
PR	16-JUN-1999; 99US-0139452.	PR 02-AUG-1999;	99US-0146386.
PR	17-JUN-1999; 99US-0139492.	PR 02-AUG-1999;	99US-0146388.
PR	18-JUN-1999; 99US-0139454.	PR 02-AUG-1999;	99US-0146389.
PR	18-JUN-1999; 99US-0139455.	PR 03-AUG-1999;	99US-0147038.
PR	18-JUN-1999; 99US-0139456.	PR 04-AUG-1999;	99US-0147204.
PR	18-JUN-1999; 99US-0139457.	PR 04-AUG-1999;	99US-0147302.
PR	18-JUN-1999; 99US-0139458.	PR 05-AUG-1999;	99US-0147192.
PR	18-JUN-1999; 99US-0139459.	PR 05-AUG-1999;	99US-0147260.
PR	18-JUN-1999; 99US-0139460.	PR 06-AUG-1999;	99US-0147303.
PR	18-JUN-1999; 99US-0139461.	PR 06-AUG-1999;	99US-0147416.
PR	18-JUN-1999; 99US-0139462.	PR 09-AUG-1999;	99US-0147493.
PR	18-JUN-1999; 99US-0139463.	PR 09-AUG-1999;	99US-0147935.
PR	18-JUN-1999; 99US-0139763.	PR 10-AUG-1999;	99US-0148171.
PR	21-JUN-1999; 99US-0139817.	PR 11-AUG-1999;	99US-0148319.
PR	22-JUN-1999; 99US-0139899.	PR 12-AUG-1999;	99US-0148341.
PR	23-JUN-1999; 99US-0140353.	PR 13-AUG-1999;	99US-0148565.
PR	23-JUN-1999; 99US-0140354.	PR 13-AUG-1999;	99US-0148684.
		PR 16-AUG-1999;	99US-0149368.
		PR 17-AUG-1999;	99US-0149175.
		PR 18-AUG-1999;	99US-0149426.
		PR 20-AUG-1999;	99US-0149722.
		PR 20-AUG-1999;	99US-0149723.
		PR 20-AUG-1999;	99US-0149929.
		PR 23-AUG-1999;	99US-0149902.
		PR 23-AUG-1999;	99US-0149930.
		PR 25-AUG-1999;	99US-0150566.
		PR 26-AUG-1999;	99US-0150884.
		PR 27-AUG-1999;	99US-0151065.
		PR 27-AUG-1999;	99US-0151066.
		PR 30-AUG-1999;	99US-0151303.
		PR 31-AUG-1999;	99US-0151438.
		PR 01-SEP-1999;	99US-0151930.

Db 34040 CCGACACCGACACCAACGCCCACTCCACACACCGACCCACACCCC 33997

RESULT 7
AAC44444
ID AAC44444 standard; DNA; 777 BP.
AC
AC AAC44444;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 42845.
DE
DE Arabidopsis thaliana
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

Query Match	2.7%	Score	78.2	DB	21	Length	58857
Best Local Similarity	62.6%	Pred. No.	9.1e-06				
Matches 122	Conservative	0	Mismatches	73	Indels	0	Gaps
QY: 2426	cqccatcgccgtcgccgctcgccgagctcgagtcacatctcgctcgccgcttcgctgcgcgt	2485					
Db: 47143	CGCTGCTGCTGCGTGGCACTGTGGCCGTGATCGGCACCTGCTCGCGCCGCGGCCGT	47084					
QY: 2486	caccatcgccgagtcgctctctggtctctcgctcaccatcgcgctcgccgagccgcttctgct	2545					
Db: 47083	CGTCTGCGCCGGGGCCGTCTGCTGCGCGGGGGCCGTGCTGCTGGCGGGGGCGTGTGCT	47024					
QY: 2546	caccagacgcgtcttctgctcaccgtctctgctcgccgagcccaacgctgctgctgcgcgg	2605					
Db: 47023	GGCCGGGGCGGTCTGCTGCGGGCCGCTGCTGCTGCGCGGGCTGCTGCTGTGACCCGG	46964					
QY: 2606	tgtcgggtgggggtga	2620					
Db: 45963	CGAGCGAGCGCAGGA	46949					

RESULT	13	
AAS84956/c		
ID	AAS84956 standard; cDNA; 651 BP.	
XX		
XX	AAS84956;	
XX	AC	
XX	DT	
XX	13-FEB-2002 (first entry)	
XX		
DE	DNA encoding novel human diagnostic protein #20760.	
DE		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
XX	Homo sapiens.	
XX	OS	
XX	WO200175067-A2.	
PN		
XX	11-OCT-2001.	
PD		
XX		
XX	30-MAR-2001; 2001WO-US08631.	
XX		
XX	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
XX	(HYSE-) HYSEQ INC.	
PA		
XX		
XX	Drmanac RT, Liu C, Tang YT;	
PI		
XX		
XX	WPI; 2001-639362/73.	
DR	P-PSDB; ABG20769.	
XX		
XX	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity -	
XX		
PS	Claim 1: SEQ ID No 20760; 103pp; English.	

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A564197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 651 BP: 152 A: 86 C: 241 G: 172 T: 0 other;

[illegible]

RESULT 14
AAS46648/c
ID AAS46648 standard; DNA; 6510 BP.
XX
XX AC AAS46648;
XX
DT 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #370.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
XX Homo sapiens.
OS
XX
XX WQ200168912-A2.
PN
XX
PD 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.
PF
XX
XX 15-MAR-2000; 2000DE-1013847.
PR
XX 06-APR-2000; 2000DE-1019058.
PR
XX 07-APR-2000; 2000DE-1019173.
PR
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043829.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2001-602752/68.
DR
XX
XX Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT

THIS PAGE BLANK (uspto)

Result No.	Score	Query			DB	ID	Description
		Match	Length				
c 1	100.8	3.5	390	4	US-09-197-649-7	Sequence 7, Appli	
c 2	80	2.8	11707	4	US-09-136-574A-1	Sequence 1, Appli	
c 3	79	2.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli	
c 4	71.2	2.5	494	4	US-09-056-556-176	Sequence 176, Appli	
c 5	70.2	2.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli	
c 6	70.2	2.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli	
c 7	70	2.4	390	4	US-09-197-649-7	Sequence 7, Appli	
c 8	68.8	2.4	18596	4	US-09-318-448-11	Sequence 11, Appli	
c 9	68.4	2.4	1931	2	US-09-130-114-2	Sequence 2, Appli	
c 10	67.8	2.4	407	4	US-09-056-556-173	Sequence 173, Appli	
c 11	67.8	2.4	23673	4	US-09-773-816-1	Sequence 1, Appli	
c 12	66.6	2.3	2293	1	US-08-604-913B-12	Sequence 12, Appli	
c 13	66.4	2.3	6416	4	US-09-136-574A-2	Sequence 2, Appli	
c 14	66.2	2.3	1294	3	US-09-025-691-2	Sequence 2, Appli	
c 15	65	2.3	3004	1	US-08-276-213-6	Sequence 6, Appli	
c 16	64.4	2.2	2064	1	US-08-343-428-1	Sequence 1, Appli	
c 17	64	2.2	985	4	US-09-056-556-182	Sequence 182, Appli	
c 18	62	2.2	1280	4	US-09-060-756-4	Sequence 4, Appli	
c 19	61.8	2.2	6085	4	US-09-029-603-4	Sequence 4, Appli	
c 20	61.6	2.1	1881	2	US-08-596-300A-4	Sequence 4, Appli	
c 21	61.6	2.1	3041	2	US-08-596-300A-5	Sequence 5, Appli	
c 22	61.6	2.1	3718	2	US-08-596-300A-6	Sequence 6, Appli	
c 23	61.6	2.1	3796	2	US-08-596-300A-3	Sequence 3, Appli	
c 24	61	2.1	1294	4	US-09-312-038-4	Sequence 4, Appli	
c 25	61	2.1	2289	4	US-09-312-038-3	Sequence 3, Appli	
c 26	60	2.1	535	4	US-09-056-556-171	Sequence 171, Appli	
c 27	60	2.1	1439	4	US-09-056-556-167	Sequence 167, Appli	

QY	2644	tcgcgcgcgggtgataatcaagatcaagccgggtttggaagtggtggaataccaggtcgtcg	2703
Db	3928526	CCGCTGTGCGCGCGCCGCGCTGCGCGCGCGCGCGCGCGCGCGGATACGCGCGTTG	3928467
QY	2704	tcggtgaatttgtcgacgggtgacggtgcggtactggttcacccggatggtggtcgtcg	2763
Db	3928466	CCGCGGTTCCGCGCGGTGTGGCCAGCCGCCGACGCGCGCGTGCGCCGCGTCCGCGCG	3928407
QY	2764	acactggtgtacaactgtgactggcgcgcatcggtgtggtggaataatccgcgcctcgttc	2823
Db	3928406	GTCCGCGCGACCCGCCGATAAAGCGCTGCCGCTGCCGCGGTTCGCCGCGCCCGCGCG	3928347
QY	2824	ggctcgtgtgaaccggcgagcccgacggcgg	2854
Db	3928346	GCCCGCGCGTTCGCGCGGTACCGCTGGCGG	3928316

RESULT 4
US-09-056-556-176/c
; Sequence 176, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Sreiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241

```
,  
,  
, ADDRESS: SEED AND BERRY LLP  
, STREET: 6300 Columbia Center, 701 Fifth Avenue  
, CITY: Seattle  
, STATE: Washington  
, COUNTRY: USA  
, ZIP: 98104-7092  
, COMPUTER READABLE FORM:  
, MEDIUM TYPE: Floppy disk  
, COMPUTER: IBM PC compatible  
, OPERATING SYSTEM: PC-DOS/MS-DOS  
, SOFTWARE: PatentIn Release #1.0, Version #1.30  
, CURRENT APPLICATION DATA:  
, APPLICATION NUMBER: US/09/056,556  
, FILING DATE: 07-APR-1998  
,
```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 176:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-176

Query Match	2.5%	Score	71.2	DB	4	Length	494
Best Local Similarity	51.9%	Pred. NO.	4.2e-07				
Matches	160	Conservative	0	Mismatches	148	Indels	0
						Gaps	0

Qy	2347	gcacccatccgagatccgcctctccgcttcgcttgtagtcgctcgagcttcgtagcagctgagccgcg	2406
Db	345	gcgcgcgcaagcgcgcgctgttcgccccgctggagccgcgcgcctatgccccccttgcgccaccg	286
Qy	2407	agccacgacccgagcagctcgcacatcgccgttcgcgcgttcgcgcagctcgagttccatctcg	2466
Db	285	aggccgcctctggccgcgcgcgcacgcgcaaaagacgcgcgttcgcgcgcgcgcgcgcgcgcacacgcg	226
Qy	2467	tcgcgcgtctcgttcgcgcgttcacacatcgcgcagtcgcgttcgctgcgtcttcgcgttcacacatgcgcg	2526

RESULT 6

```

Db      225  CCGTTGCCGCGCCACCGGCCACGGTCCGACAGTAGCCGCGCGCTTGGGGCCGCCCTGA 166
Qy      2527  tcgcgagcccgcttcgtcaaccgagcccgctcttcgtcaccgctcttcgtcgccgagccca 2586
Db      165  CCGCGCTGCGCGCGAGCGCCCTTGCGCGCGGAAAGCGCTGCCGCGCGCGCGCGCG 106
Qy      2587  acgcgctcgtcgtccgcggtgcggtgggtgaaggtgcagtaagaataatgattcg 2646
Db      105  GCGCGCGCGCCACCGCGCGCTCCGCGCTTGGGTACGGGAACCTGTTGCCGCCGACGCCGTCA 46
Qy      2647  gcgcgcggg 2654
Db      45  CCGCGGTG 38

RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

[illegible]

RESULT 6

Patent No. 5336623

GENERAL INFORMATION:

APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert

APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: USA

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/276,213

APPLICATION NUMBER: US/08/276,213

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3004 base pairs


```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-276-213-6
```

```
Query Match      2.3%; Score 66; DB 1; Length 3004;
Best Local Similarity 59.7%; Pred. No. 8.5e-06;
Matches 111; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 2429 catcgccgtcgccgtcgccgagctcgagtcacatcctcgctcgccgtctccgtcgccgtcac 2488
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2025 CGTCTGCATCGCCTAGCAGTCAACCGTCCCGTCCGTGTCGCCCTCTCCGTCCGCCGCC 2084

Qy 2489 catcgccgagtcgctctcggtctcctcgtaaccatcgccgtcgccgagcccgctcttcgtcac 2548
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2085 CGTCGGCGAGTCGGACGCCGCGCTACTCGAGCGCGACAGCCGAGCCGCGACGCCAACGC 2144

Qy 2549 cgagccgtcttcgtaccgtcttcgtcgccgagcccaagcccgctcgctcgccggtgt 2608
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2145 TGACCCCTACTGCTACGCCCCACGCCCGCAAGCCCGACCGCTACCCGACGCCGCT 2204

Qy 2609 cgggtg 2614
    |||
Db 2205 CCGGAG 2210
```

Search completed: July 2, 2002, 16:53:17
 Job time: 27704 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2002, 09:11:33 ; Search time 2155.46 Seconds
(without alignments)
17964.965 Million cell updates/sec

Title: US-09-917-376-2

Perfect score: 2869
Sequence: 1 atgagatcgttcggagaacat.....ggcggaacacacacacgcagn 2869

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_estt2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.4	3.3	559	12 BH306809	BH306809 CH230-100
2	92.2	3.2	464	12 A2928926	A2928926 479.dif19
3	90	3.1	546	10 BF072664	BF072664 NCSM3H7T3
4	84.8	3.0	306	12 CNS07HOW	AL610466 Anopheles
5	81.6	2.8	236	9 AA720369	AA720369 33562 Lam
6	81.2	2.8	478	10 BE496065	BE496065 WHEI261_G
7	81.2	2.8	523	10 BF145280	BF145280 WHEI834_f
8	81	2.8	1315	9 AV752548	AV752548 AV752548
9	80	2.8	565	12 AZ386138	AZ386138 LM0144P14
10	79.6	2.8	873	9 BE217030	BE217030 EST0425 T
11	78.4	2.7	527	12 A2297141	A2297141 RPCI-23-1
12	75	2.6	1057	12 CNS053F1	AL319366 Tetraodon
13	74.6	2.6	401	10 BE499920	BE499920 WHE0976.F
14	74.2	2.6	321	10 BG874723	BG874723 RC1-BT025
15	72	2.5	535	10 BE494335	BE494335 WHEI253_D
16	72	2.5	1208	9 AL514927	AL514927 AL514927
17	71.6	2.5	460	12 CNS036QU	AL230367 Tetraodon

18	71.6	2.5	522	9 AI601018	AI601018 486096B12
19	71.6	2.5	606	9 AI603748	AI603748 486096B12
20	71.2	2.5	587	10 C79507	C79507 C79507 Mous
21	71	2.5	259	12 AZ374273	AZ374273 IM0126G21
22	70.2	2.4	809	12 CNS02CFH	AL191078 Tetraodon
23	70	2.4	487	10 BE494661	BE494661 WHEI257_G
24	70	2.4	490	10 BE494611	BE494611 WHEI255_C
25	69.8	2.4	859	10 BF627594	BF627594 HVSMEB000
26	69.2	2.4	439	9 AW496772	AW496772 ESTPM077
27	69.2	2.4	570	10 BG366070	BG366070 HVSMEI000
28	69.2	2.4	696	9 AL506972	AL506972 AL506972
29	69.2	2.4	710	9 AU163439	AU163439 AU163439
30	68.6	2.4	365	12 L2644X	AL160994 Leishmani
31	68.4	2.4	557	10 BF145396	BF145396 WHEI841-I
32	68.2	2.4	279	9 AW496745	AW496745 ESTPM041
33	68.2	2.4	462	9 AU197994	AU197994 AU197994
34	68.2	2.4	623	10 BE495160	BE495160 WHEI274_B
35	68	2.4	373	10 BF429394	BF429394 WHEI804_F
36	67.8	2.4	584	10 BF255396	BF255396 HVSMEF000
37	67.8	2.4	955	12 CNS02YA4	AL219397 Tetraodon
38	67.6	2.4	554	9 AW677991	AW677991 WSI_12_HO
39	67.2	2.3	427	10 BE494411	BE494411 WHEI254_E
40	67.2	2.3	450	9 AU198202	AU198202 AU198202
41	67.2	2.3	829	12 CNS035GN	AL228704 Tetraodon
42	67	2.3	504	10 BM381428	BM381428 MES7534-F
43	67	2.3	651	10 BE419325	BE419325 WWR06.G10
44	66.8	2.3	1009	12 CNS04065	AL268502 Tetraodon
45	66.6	2.3	458	10 BE587118	BE587118 WHE0517_D

ALIGNMENTS

RESULT 1

BH306809
LOCUS BH306809 559 bp DNA linear GSS 30-NOV-2001
DEFINITION CH230-100P7, TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH306809
VERSION BH306809.1 GI:17219217
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 559)

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-100P7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orering.information.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 100 row: P column: 7
Seq primer: 37
Class: BAC ends.
Location/Qualifiers
1..559
/organism="Rattus norvegicus"
/strain="BN/SsNHsd/MCW"


```
/db_xref="taxon:10116"
/clone="CH230-100P7"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
Pieter de Jong"
BASE COUNT      72 a   232 c   65 g   190 t
ORIGIN

Query Match      3.3%; Score 93.4; DB 12; Length 559;
Best Local Similarity 65.5%; Pred. No. 7.6e-05;
Matches 167; Conservative 0; Mismatches 86; Indels 2; Gaps 2;

QY 2350 ccgtccgagatcgccgtctccgttcggtgagtcctcggttcgagcctgagcccgagc 2409
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 268 ccgtctctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtct 327
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2410 ccgagcccgagcagtcgccatcgccgtccggtccgagcgtcgag-tccatctcgtc 2468
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 328 ccgtctccgtctccgtctccgtctccgtctccgtctccgtctccgtctccgtctcc 387
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2469 gccgtctccgtccgtccatccatccgagtcggtctccgtctccgtccatccatccgcgc 2528
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 388 tcgctctccgtctccgtctccgtctccgtctccgtctccgtctccgtctccgtctcc 446
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2529 gcgagcccgcttcgtccacgcgcgcgttcgttcacgcgttcgttcgagcccgccaac 2588
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 447 tcctctccgtctccgtctccgtctccgtctccgtctccgtctccgtctccgtctcc 506
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2589 gccgtctcgtcgcc 2603
||||| |||||
Db 507 tcgctctccgtctcc 521

RESULT 2
AZ928926/c
LOCUS
DEFINITION
479.dif19g12.sl Saccharomyces kluyveri DNA sequence.
GSS 01-APR-2001
ACCESSION
AZ928926
VERSION
AZ928926.1 GI:13499832
KEYWORDS
GSS.
SOURCE
Saccharomyces kluyveri.
ORGANISM
Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 464)
Clifton,P.F., Hillier,L.W., Fulton,L.W., Graves,T., Miner,T., Gish
.W.R., Waterston,R.H. and Johnston,M.
Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
Unpublished (2001)
Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
FEATURES
Location/Qualifiers
1..464
/organism="Saccharomyces kluyveri"
/strain="NRRL Y-12651 (CBS 3082)"
/db_xref="taxon:4934"
/clone="479.dif19g12.sl"
/clone_lib="Saccharomyces kluyveri"
/note="Random genomic sequence"
BASE COUNT      119 a   69 c   163 g   113 t
ORIGIN

Query Match      3.2%; Score 92.2; DB 12; Length 464;
Best Local Similarity 68.6%; Pred. No. 0.00011;
Matches 127; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2425 tcgccatcgccgtcgccgtcgagtcgagtccttccttcgttcgcttcgcttcgctgc 2484
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 323 TCGCCATCACCATCGCCATCGCCATCACCATCACCATCACCAGCACCAGCACCAGCACC 264
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2485 tcaccatcgccgagtcgcttcggttcggttcggttcggttcggttcggttcggttcggt 2544
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 263 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCAC 204
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2545 tcaccgagccgcttcgctacgcttcgttcgagccgagccgagccgagccgagccgag 2604
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 203 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCAC 144
||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
QY 2605 gtgtc 2609
|||||
Db 143 TCGTC 139

RESULT 3
BF072664
LOCUS
DEFINITION
NCSM3H7H3 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5'
similar to avicelase III, Aspergillus aculeatus, mRNA sequence.
ACCESSION
BF072664
VERSION
BF072664.1 GI:10866169
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE
1 (bases 1 to 546)
Neelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K., Miller,R.,
Ortega,J., Pavlova,I., Perea,J., Todisco,S., Trujillo,R.,
Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S. and Natvig
,D.O.
Expressed sequences from conidial, mycelial, and sexual stages of
Neurospora crassa
Fungal Genet. Biol. 21, 348-363 (1997)
97435549
Contact: Natvig,D.O./Nelson,M.A.
Department of Biology
University of New Mexico
Catteter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngp@biology.unm.edu.
FEATURES
Location/Qualifiers
1..546
/organism="Neurospora crassa"
/strain="74-OR23-IV A (FGSC 2489)"
/db_xref="taxon:5141"
/clone="SM3H7"
/clone_lib="Subtracted Mycelial"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="Mycelium"
/lab_host="E. coli"
/note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
XhoI; 2% sucrose for 24 hours. cDNA directionally cloned
into pBluescript SK(-) using the Uni-ZAP XR vector system
(Stratagene, La Jolla, CA). Previously identified highly
expressed clones were subtracted from this library."
BASE COUNT      97 a   205 c   149 g   95 t
ORIGIN
```



```
Query Match          3.1%; Score 90; DB 10; Length 546;
Best Local Similarity 52.2%; Pred. No. 0.00024;
Matches 266; Conservative 0; Mismatches 220; Indels 24; Gaps 2;

QY 1266 gaagtcgctggtgatgaagcagtggaatcgatcgcttcaactctgactgctgct 1325
DB 52 GCACCTCGGCTGGATGATCGAATCCCTCGAGATCAACCCCTCGACAGCACCATTTGGCT 111

QY 1326 ctacgaaacaggcgcagctgtgtacgcaacaaatgatctcaacgaagtgggaactccgcgg 1385
DB 112 CTACGGCACCGGCTCTACTGTTTACGGCGGCCACGACCTGACCAAGTGGGACACCGTCCA 171

QY 1386 ccagattcatatcgcccgatgtcaaaagattggagagacgcggttaaacgattctcat 1445
DB 172 CAAGTGCACCATCCAGTCTTGCGCTTGCGATCGAAGAAATGGCTGTCTCTCGGTCTCGC 231

QY 1446 cagccgcgcgtctggcgcccgctcatcgctctcggagacgtcgcggtctca---- 1501
DB 232 CTCGGCCCCGGTGGCTCGGAGCTTCTCGGGCGCGTCTCGGTGACGACTGGCGCTTACCTT 291

QY 1502 -----cccaagcgcagcttactgcgtgcacatcgacgatcttcacgtcacccggtgtcac 1556
DB 292 CAAGTCCAGTCCGACCTCGCGACCTCTCCCAAGACACCCCTGGATGACGCCCCAGTGGC 351

QY 1557 gaccgcaccagcgtcgactatcggaattgaatccgtcgatcatcgttcgcgtgaaag 1616
DB 352 CAGCTCGGGAGCTGCTACGCCGGCAACAAGCCGCCCAACGCTGGTGGCGATCGG--- 408

QY 1617 ttctgatccatcgagcacaacccgacagcagcgctcgtcttcacagacgcgcgcaaa 1676
DB 409 -----CTCGGGCTCCGGCGCGCAGCAGAGTGGCGGTATCTCTCGGACGCGCGCGC 456

QY 1677 gaactggttccaaggcagcgaacctggcggggtgagacggcgccacgcgtccgcacatc 1736
DB 457 CTCGTGGCATGCTCACAACGGCACCCACACACCAAGAGCAGCGGACGCTCGCCTACTC 516

QY 1737 ggcgcagcgtctcgtttcgtctggctcc 1766
DB 517 TGCCGAGCGCGACCATCATGCTCTGGTCTTC 546

RESULT 4
CNS07HOW
LOCUS
DEFINITION
  CNS07HOW 306 bp DNA linear GSS 02-OCT-2001
  Anopheles gambiae GSS SP6 end of clone 24A06 of library NotreDame1
  from strain PEST of Anopheles gambiae (African malaria mosquito),
  genomic survey sequence.
ACCESSION
  AL610466
VERSION
  AL610466.1 GI:15916651
KEYWORDS
  GSS.
SOURCE
  African malaria mosquito.
  Anopheles gambiae
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
  Culicoidae; Anopheles.
REFERENCE
  1 (bases 1 to 306)
  Genoscope.
  Direct Submission
  Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
REFERENCE
  2 (bases 1 to 306)
  Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H.
  Direct Submission
  Submitted (01-OCT-2001) BBMI, Institut Pasteur, 25, rue du Dr.
  Roux, Paris 75015, France
  This clone is from an A. gambiae BAC library provided by F.H.
  Collins and sequenced by Genoscope in collaboration with the
  Laboratory of Biochem. and Biol. Molec. of Insects, Institut
  Pasteur.
  Location/Qualifiers
    1..306
    /organism="Anopheles gambiae"
```

```
/strain="PEST"
/db_xref="taxon:7165"
/clone="24A06"
/clone_lib="NotreDame1"
/note="end : SP6"
BASE COUNT      65 a 110 c 63 g 68 t
ORIGIN

Query Match          3.0%; Score 84.8; DB 12; Length 306;
Best Local Similarity 70.6%; Pred. No. 0.0014;
Matches 113; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2425 tqccatcgccgtcgccgtcgagctcgagtcacatctcgtcgccgtctcgttcgcgcg 2484
DB 120 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCG 179

QY 2485 tcaccatcgccagtcgctcgtcgttcgctcaccatcgcgctcgcgagccgctcttcg 2544
DB 180 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTCACCG 239

QY 2545 tcaccgagccgctcttcgtcaccgcttcgttcgctcccgagcc 2584
DB 240 TCACCGTCACCGTCACCGTCACCGTCACCGTCACCGTTACCGTGCC 279

RESULT 5
AA720369
LOCUS
DEFINITION
  AA720369 236 bp mRNA linear EST 30-DEC-1997
  33562 Lambda-PRL2 Arabidopsis thaliana cDNA clone 171N14T7, mRNA
  sequence.
ACCESSION
  AA720369
VERSION
  AA720369.1 GI:2733979
KEYWORDS
  EST.
SOURCE
  thale cress.
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
  1 (bases 1 to 236)
  Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
  ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
  ,E. and Somerville,C.
  Genes galore: a summary of methods for accessing results from
  large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  Plant Physiol. 106, 1241-1255 (1994)
  95148729
  Contact: Thomas Newman
  MSU-DOE Plant Research Laboratory
  Michigan State University
  MSU-DOE-PRL, Michigan State University, Plant Biology Bldg.,E.
  Lansing,MI
  Tel: 517-353-0854
  Fax: 517-353-9168
  Email: 22313tcn@bm.cl.msu.edu
  Seq primer: T7 dye primer.
  Location/Qualifiers
    1..236
    /organism="Arabidopsis thaliana"
    /strain="var columbia"
    /db_xref="taxon:3702"
    /clone="171N14T7"
    /clone_lib="Lambda-PRL2"
    /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
    Lambda PRL2 is a cDNA library derived from equal
    quantities of 4 pools of mRNA. The mRNA sources were 1) 7
    day germinated etiolated seedlings; 2) tissue culture
    grown roots; 3) staged plants half with 24 hour light
    cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
    same plants as 3 but aerial tissue (stems, flowers and
    siliques. The vector is BRL's lambda Zip-Lox. The cDNA
    inserts were directionally cloned with Sal-Not arms using
    oligo dT primed cDNA."
```


greenhouse. Insects were harvested and pooled for meiosis to late meiosis. The tissue, total RNA, and poly(A) RNA were prepared (Butler, Ross and Gustafson) at

BASE COUNT	408 a	403 c	224 g	280 t
ORIGIN	XhoI"			
Query Match	2.8%; Score 81; DB 9; Length 1315;			
Best Local Similarity	65.4%; Pred. No. 0.0055;			

Query Match 2.0%; Score 81; DB 9; Length 1319;
Best Local Similarity 65.4%; Pred. No. 0.0055;
Matches 134: Conservative 0; Mismatches 70; Indels 1; Gaps 1;

[illegible][illegible][illegible][illegible]

		9
RESULT	AZ386138	
LOCUS	AZ386138	DNA linear GSS 02-OCT-2000
		565 bp

DEFINITION
clone U4GC1M0144P14 R, DNA sequence.
ACCESSION
AZ386138
VERSION
AZ386138.1 GI:10499754

GSS.
 AIRWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
1. (bases 1 to 565)
Mammalia; Eutheria
Barber, M., Beacom, T., Duval, B., Hamll, C.,
Kodner, A., Meenen, E., Pedersen, T., Reilly,
S., Mahmoud, M., Longacre, S., Longacre,
Dunn, D., Aoyagi, A.,
Islam, H., Longacre, S., Mahmoud, M.,
1. (bases 1 to 565)
Mammalia; Eutheria
Barber, M., Beacom, T., Duval, B., Hamll, C.,
Kodner, A., Meenen, E., Pedersen, T., Reilly,
S., Mahmoud, M., Longacre, S., Longacre,
Dunn, D., Aoyagi, A.,
Islam, H., Longacre, S., Mahmoud, M.,

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
M., ROSE, R., ROSE, R., and Wright, D., Weiss, R.,
Stokes, R., Lingey, A., von Niederhausern, A.,

JOURNAL COMMENT

unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Genome Center
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
plate: 0144 row: P column: 14
Seq primer: CACACACGGAACACGCTATACCC

```

FEATURES
SOURCE
1
Location/Qualifiers
525
High quality sequence stop: 565.
Class: plasmid ends

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/ncbi_refseq="NM_001144911"

```

```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

```

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA

© 2007 Pearson Education, Inc. All rights reserved. Printed in the United States of America. This publication is protected by copyright. Any unauthorized reproduction or distribution of this work in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from Pearson Education, Inc., is prohibited. This publication may be reproduced in whole or in part for personal or internal reference use only on the basis of written permission from Pearson Education, Inc. For more information, contact Pearson Education, Inc., 501 Boylston Street, Boston, MA 02116.

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114 (gb)AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

BASE COUNT      121 a  175 c  112 g  157 t
ORIGIN

Query Match      2.8%; Score 80; DB 12; Length 565;
Best Local Similarity 63.5%; Pred. No. 0.0076;
Matches 122; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 2401 agccagagccgagccgagcagctccatccgctccgctccgagctcgagtcga 2460
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 ATCTGAGCTCGTGAAGAAAGCTTATCCCTCTCTCCGCTCCGTTTCCTCTCCGTC 428

QY 2461 tcctcgtcgcgtctcgtcgcgtccatccgctccgagctcgtctcgtccgta 2520
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 TCTCCGCTCCGCTCCCTCTCCGCTCCCTCTCCGCTCCCTCTCCGCTCCCTCTCCG 488

QY 2521 tcggcgtcgcgagccgcttctcgtccgagccgctcgtcgtccgagctcgtc 2580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 489 TCTCCCTCTCCGCTCCCTCTCCGCTCCCTCTCCGCTCCCTCTCCGCTCCGTC 548

QY 2581 agcccaacgcgc 2592
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 549 TCTCTCCCTCCCTG 560

RESULT 10
BE217030
LOCUS      BE217030      873 bp      mRNA      linear      EST 03-JUL-2000
DEFINITION      EST0425 Triticum aestivum Lambda zap Triticum aestivum cDNA clone
                JAL_5A_H09_T3 5', mRNA sequence.
ACCESSION      BE217030
VERSION        BE217030.1 GI:8904570
KEYWORDS       EST.
SOURCE         bread wheat.
ORGANISM       Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 873)
Anderson,J.M., Williams,C.E. and Goodwin,S.B.
Analysis of an EST database reveals a probable Cf2 resistance gene
homolog in wheat
Unpublished (2000)
Contact: Anderson, J.M.
Crop Production & Pest Control Research Unit
USDA-ARS
1150 Lilly Hall, West Lafayette, IN 47907, USA
Tel: 765-494-5565
Fax: 765-496-2926
Email: janderson@purdue.edu
Seq primer: T3
High quality sequence stop: 873.
Location/Qualifiers
1..873
/organism="Triticum aestivum"
/strain="P29"
/db_xref="taxon:4565"
/clone="JAL_5A_H09_T3"
/clone_lib="Triticum aestivum Lambda Zap"

0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g14732114 (gb)AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      119 a  344 c  276 g  133 t  1 others
ORIGIN

/tissue_type="leaf"
/dev_stage="9 day old seedlings"
133 t  1 others

Query Match      2.8%; Score 79.6; DB 9; Length 873;
Best Local Similarity 57.1%; Pred. No. 0.0088;
Matches 145; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 2363 cgtctcgttcggtgagtcgcttcgctccgagctcgtcgtcgtcgtcgtcgtc 2422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 CATCCGCGCGCGGACACCTCCCTCGCGCCCTCGCGCCCTCGCGCCCTCGAGTG 276

QY 2423 gctcgcacgcgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgc 2482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 277 GCTCGCGCGGAGCTCGCGCGCTCAAGCTCGACCGCGCGGCTCTCTCTCTCTCG 336

QY 2483 cgtcaccatcgcgcgttcgtctcgtctcgtctcgtctcgtctcgtctcgtctc 2542
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 CGTCCGCGCTCGCGGACGACGACCGCGCGCGCTCGCTCGCTCGCTCGCTCTC 396

QY 2543 cgtcacgagccgctcttcgtctcgtctcgtctcgtctcgtctcgtctcgtcgc 2602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 397 CGCCCCGACCGCGCTCGACGCGCTCGCGCGCTCGCTCGCGCTCTCGCGCCACTC 456

QY 2603 cgtgtcgtcgtggtggtg 2616
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 CGTGGCGCGGACGCG 470

RESULT 11
AZ297141/c
LOCUS      AZ297141/c      527 bp      DNA      linear      GSS 27-JUL-2000
DEFINITION      RPCI-23-164L13-TV RPCI-23 Mus musculus genomic clone RPCI-23-164L13
                DNA sequence.
ACCESSION      AZ297141
VERSION        AZ297141.1 GI:9538926
KEYWORDS       GSS.
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 527)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-164L13.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@delong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 164 row: L column: 13
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-164L13"
/clone_lib="RPCI-23"

FEATURES
source
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-164L13"
/clone_lib="RPCI-23"

```


THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:39 ; Search time 101.03 Seconds
(without alignments)
813.566 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATQPTWSNVAIGGGFVD.....YGTNGRGIVYDGGAPSG 740

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	4.0	1684	AA19498	Bacterial amylase
2	159.5	4.0	1291	AAW59912	Amino acid sequenc
3	158.5	3.9	774	AAW35390	Flavobacterium odo
4	152.5	3.8	1751	AA113493	Truncated cellulase
5	152	3.8	745	AAAP90615	Polypeptide with i
6	152	3.8	3472	AAAP90913	Cenarchaeum symbio
7	151.5	3.8	776	AAAY90916	Isoamylase amino a
8	150.5	3.7	2367	AAU31850	Novel human secret
9	148.5	3.7	750	AAAP94419	Primary transcript
10	148	3.7	1045	ABG08332	Novel human diagno
11	148	3.7	1532	AAAB40945	Human ORFX ORF709

12	146	3.6	1749	22	AA172314	Glutamic acid rich
13	144.5	3.6	1472	22	AA159423	Drosophila melanog
14	143.5	3.6	1532	22	AAU71956	Human bone marrow
15	143	3.5	894	15	AAAR47578	Flocculation prote
16	143	3.5	894	15	AAAR58754	S. cerevisiae PLO1
17	142.5	3.5	995	15	AAAR60812	Agarase 0107 from
18	142	3.5	925	22	AAAG92740	C glutamicum prote
19	141	3.5	754	21	AAAB35987	Sorbitol dehydroge
20	140.5	3.5	1000	21	AAAY58914	Fumonisin esterase
21	140.5	3.5	1000	21	AAAY68850	A fusion of aminop
22	140.5	3.5	1403	22	AAAB60540	Drosophila melanog
23	139	3.4	3118	22	AAU27790	Human full-length
24	137.5	3.4	1039	22	AAAB61734	Drosophila melanog
25	137.5	3.4	2464	22	AAAB63174	Drosophila melanog
26	137	3.4	911	22	AAU56103	Protonibacterium
27	136.5	3.4	1225	20	AAAY34538	Porphorymonas ging
28	136.5	3.4	1226	20	AAAY34537	Porphorymonas ging
29	136.5	3.4	1251	20	AAAY34404	P. gingivalis haem
30	136	3.4	2628	17	AAAR60300	Haemagglutinin pro
31	136	3.4	2628	19	AAAW69488	Fumonisin esterase
32	135.5	3.4	1205	21	AAAY58916	A fusion of aminop
33	135.5	3.4	1205	21	AAAY68852	Drosophila melanog
34	135	3.3	823	22	AAAB71484	Drosophila melanog
35	134.5	3.3	1805	22	AAAB65262	Mannanase. Cellul
36	134	3.3	1010	22	AAAB49442	Human polypeptide
37	133.5	3.3	530	22	AAO13658	Amino acid sequenc
38	133.5	3.3	1429	21	AAAY93606	Escherichia coli p
39	133.5	3.3	1778	22	AAAB52677	Bankia gouldi endo
40	133	3.3	875	19	AAAB34987	Drosophila melanog
41	133	3.3	1357	22	AAAB70775	Human wound healin
42	133	3.3	1712	22	AAAB44566	Human polypeptide
43	133	3.3	2058	22	AAAB97070	Soluble chitinase.
44	132.5	3.3	866	17	AAW02159	Vibrio furnissii e
45	132.5	3.3	866	21	AAAY52307	

ALIGNMENTS

RESULT 1
AA19498
ID AA19498 standard; Protein; 1684 AA.

AC AA19498;

XX 25-FEB-1992 (first entry)

DT Bacterial amylase A-180.

DE Maltopentose; G5; starch hydrolysis.

KW EP459385-A.

XX 04-DEC-1991.

XX 28-MAY-1991; 91EP-0108669.

XX 31-MAY-1990; 90DE-4017595.

XX (CONE) CONSORT ELEKTROCHEM IND.

XX Schmid G, Candussio A, Bock A;

XX WPI; 1991-355676/49.

DR N-PSDB; AAQ13939.

PT New bacterial amylase, A-180 for malto:pentose prodn. - by

PT hydrolysis of starch, providing high yield and modifiable for

PT secretion from host cells

XX Disclosure; Page 7; 21pp; German.

CC The amino acid sequence is that of bacterial amylase A-180 which is

QY 180 TT---GVQSDIOGVV-----WVAFDKSSSLGOASKTIFVGVADPNPNVFWSDGGA 228
Db 396 ltlldganynitsiviklnpssiwaaartqtiqlgh-----dqnttfsnlvsak 445
QY 229 TQOAVPCA-PTGIPHPKGVDPVNNHLYIATSNCTGPGYDSSGDVWKF-----275
Db 446 sysfdpasntvtip--vtatvkrlqlnitsnsgp---agvvaefqvgtpapnptdi 498
QY 276 SVTSGTWTRISPVSDTDANDYFGYSGGLID---RQHPNFMVATOISWMPDIIIFRSTD 332
Db 499 tltgmswspspv-etd-----aitlnatvkngnasaatvtnfylnnelagsap 548
QY 333 GGATWTRINDWTSYPNRSRLRYLVIDIAEPWLTGQVQPNPPVPSKLGWDEMAIDPFNS 392
Db 549 vaal-----aagasaatvplnvgaaktaatyavgak-----vdesnavielne 599
QY 393 DRMLYGTGATL---YATNDL---TKW-----DSGGQIHIA 421
Db 590 snnsytnpaslvvapyvsssdlygtvstptpiannavsfvnlknqgtiasaggshgvt 649
QY 422 PMYKGLEETAVNDLISPPSG--APLISA-----LGLDGGFT-----HADVTAVP---463
Db 650 vviknasgstvtfgsgysglapgasvnltpgtwtaaagsyvtvatvaadanelpikq 709
QY 464 ---STIFSPVTTCTSDVYAEALNP-----SIIVRAGSFD-----PSSQPNDRHVAF 507
Db 710 annantasilvysargaspysrydtedatlgggatiksaaptfdqaltaseatqglyaal 769
QY 508 STDGG-KNMF--OGSEPGGVTTGGTVAASADGSRFVWAPGDPQOPVYAVGFG-----557
Db 770 psgsyqlqwtvrgggagvmtfmpdsadgm-----glngslvnyngtkvktvsit 823
QY 558 --NSWAASQG-VPANAQ-----IRSDRVNPKTFYALSNGTFFYR-----STDGGVTF- 600
Db 824 syyswyfsgdmgdpapsagrpfrfdevhkwldtplkpgdtirigknngdsleygvdfi 883
QY 601 --QPVAAGL--PSSGAVGMFHAVP--GRE-----GDLWLAASSGLYHSTN 640
Db 884 elepvpaaisrpansvsvtdygvapndgqddltafkaavnaavasdkilyipegtfhl-- 941
QY 641 GGSWSAITGVSSAV-----NVGFGKSAFGSSYPAVFVVGTTG-----678
Db 942 -gmwe-igsvsmidhittagagiytniqftnanpasggislrirtgklfnsynvlnsn 999
QY 679 -----GVTGAYRS--DDCGF-----TWVLINDDQHOYGNWGOAITGDHAN 716
Db 1000 lrsrygnavykgfmdnfgtnsvirdvww-----ehfecgfw-----vgdygh 1042

RESULT 3

AAW35390
ID AAW35390 standard; Protein; 774 AA.

XX AC AAW35390;

XX DT 09-MAR-1998 (first entry)

XX TT Flavobacterium odoratum isoamylase.

DE DE Isomylase; industrial enzyme.

KW OS Flavobacterium odoratum.

XX XX Key Location/Qualifiers

FT Peptide 1..33

FT Peptide /label= sig_peptide

FT Peptide 34..774

FT Peptide /label= mat_peptide

XX JP08023981-A.

XX PN 30-JAN-1996.

XX PD

XX

PF 20-JUL-1994; 94JP-0167267.

XX 20-JUL-1994; 94JP-0167267.

PR (NISO) NIPPON SHOKUHN KAKO KK.

XX WPI: 1998-021932/03.

DR N-PSDB; AAT95400.

XX Flavobacterium odoratum isoamylase gene - useful in industrial

PT processes

XX Claim 1; Page 6; 8pp; Japanese.

XX The present sequence is Flavobacterium odoratum isomylase,

CC which can be used as an industrial enzyme.

XX Sequence 774 AA;

SQ

Query Match 3.9%; Score 158.5; DB 19; Length 774;
Best Local Similarity 18.8%; Pred. No. 0.016;
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;

QY 76 INTNKKWAA-----VGNMTNSWDPNDGAILRSSDOGATWITPL 114

Db 35 lnpnkigaaydatkanvtfkvyskatrielyllystatgsaekakymtctnsgglswsvt-i 93

QY 115 PFKLGNGMPCRGMLAVDPNNDNILYFGAPSGKGLWRSTDGATWSOMTNPDPVGTYI 174

Db 94 p---tstlsggglg-----gtlyygyrawgnpw---pynaswtkgssl-----130

QY 175 ANPTDTTGYQSDIQG-----VWVAFDKSSSLGOASKTIFVGVADPNNP 219

Db 131 -----gfisdvaagnrfnpnkllsdpyalelshdpttat-----tngs 170

QY 220 VFWSRDGGATWQAV---PGAPTGF-----PHKGVFDPV---NHLVLIASNTG 262

Db 171 Iyas---gatyrynldgssapkgivlagdtqatgkptralkddvvyeahvrgltmndts 227

QY 263 -----GPVDGSSGDVWKFSTVSGTWTRISPV-----PSTDTANDYFGYSGTLI 305

Db 228 itaayrgtykgaglkkaalaaalgvtaleflpvqetqndtndpsstsgdnywymtlny 287

QY 306 ---DRQ-----HPNTIMVATOISWMPDIIIFRSTDGATWTRIW 341

Db 288 fapdriryaydktpgpgptrefkemvkafhdnngikvlvd-----vvynhtgeggawspt- 339

QY 342 DWTSYFNRSRLR-----YVLDISAE-PWLTFCGVQPNPPVPS-----PKLGWDEAM 395

Db 340 dkttnyitfrgldnptyysltadfqnswdntgvggnyntnrtiaqlnlvdsiaywrkl 399

QY 386 AIDPFNSDRMLYGTGATLYATNDLTAKWDSGGQIHIAPMVKGLEETAVNDLIS--pp----439

Db 400 gvdgyrfdlasvlgnsqchqcfnfkmdag-----nalnriavaelprpat 445

QY 440 --SCAPLIS---ALG-----DLGCF-----THADVTAVPSTIFTSPVTTGTSDVYA 481

Db 446 gsggvdlaepwaignsyqvggfpsgwaewngayrdvvrqapnklgsvaitg-----499

QY 482 ELNPSIIVR-AGSFDPSQPNDRHVAFSTDGKKNWFGQ---SEPGGVTGTGTVAAASADGS 537

Db 500 ----qmatrifagssd-----lygddgrkphwsvnftahdgtfklkdllyscsnkn 545

QY 538 RFVW--APGDPQOPVYAVVCFGNSW-----AASQ-----GVP-----ANAQ 571

Db 546 nqvwpypgsdggg-----dnnswdqgggiaqdqrkaarnngmalmmlsagvniygdea 599

QY 572 IRS-----DRVNPKTFFY-----ALSNGTFFRSTD-G 596

Db 600 lrsmcnncnnpynldssanwlnwrttdqnnfqsfkamiafrkaphpalrpanfyssvdmn 659

QY 597 GVTFQPVAAGLPSSGAVGVNMF-----HAYPGK-EGDLWLAASSGLYHSTNGSSWSAIT 649
 Db 660 gnmeglrwfkpgdgvadatyfdannhaiaawridsefgdtasaiyvahn-----awsa-- 714
 QY 650 GVSSAVNVGKSGAPSSYPFAVFGTIGVGTGAYRSDDCGTTWLLINDDQHOYGNWGQ 708
 Db 715 -----qvnftlpwpgagkswrvtdtcgwaegasqvqapgs-alvggentaygicgr 766

RESULT 4
 AAY13493
 ID AAY13493 standard; Protein; 1751 AA.
 AC AAY13493;
 XX 30-JUL-1999 (first entry)
 XX Truncated cellulases comprising amino acid sequence.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing.
 XX Unidentified.

OS EP921188-A2.
 XX 09-JUN-1999.
 XX 15-SEP-1998; 98EP-0810919.
 XX 19-SEP-1997; 97US-0932571.
 XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 PI Gibbs MD, Morgan H, Williams DP;
 DR WPI; 1999-315403/27.
 DR N-PSDB; AAX55662.

PT New truncated cellulase proteins, useful in detergents and for
 PT producing 'stonewashed' denim

XX Claim 7; Page 37-41; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481. Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 1751 AA;

Query Match 3.8%; Score 152.5; DB 20; Length 1751;
 Best Local Similarity 18.7%; Pred. No. 0.12;
 Matches 184; Conservative 101; Mismatches 349; Indels 351; Gaps 49;

QY 6 YTWSNVAIGGGFVDGIVFNEGAPGILYVRTD-----IGMYRW-----D 45

Db 233 ylaatgyfysggfidlgw---aavwlyiatndssyltkaeelmseyangintwtqcwdd 289
 QY 46 AANGRIWIPLDVGVNWNWNGVVSIAADPINTNKWAAVGM-YTNSWDPNDGAIIRSSD 104
 Db 290 vrygtlmlakitgkel--ykgavernldhwtidrtypkgmayltgw---gslyrat 343
 QY 105 QG-----ATWQ-----ITPLPFKLGNNMPGR-----GMGERLAVDPNDNDI 140
 Db 344 aafiacvyadwsgcdsnkktkylnfaksqidyalgst--grsfvvgfgtnypqphhnrn- 400
 QY 141 LYFGAPSGKGLWRSTDGATWSQMTNFPD-----VGTYIANPTDTTGYQSDIQGVW-- 192
 Db 401 -----ahsswansmkipeyhrhilyalvggpgsdsyndditdyvqne 444
 QY 193 VAFDKSSSLGQASKTIFVGVADPNPNVFWSDGGATQAVPCAPTGFPHKGVFPDVNH 252
 Db 445 vacydnagivgalak-----myqlyggepid--dfkaletptnd 481
 QY 253 VLYIAT--SNTGGPYDGSQDVNMFVSTGTRISPVSTPTAN-DYF-----GY 300
 Db 482 eifveskfgnsqgp---nytevisylyntgw---pprvtdklsfkyfidltellqagy 534
 QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIWWTSPNRSRLRYL-DISA 359
 Db 535 s-----pdvvkvd-----yylegklsqgyvwd---knriyyvlvdfsg 572
 QY 360 EPWLTFGVOPNPPVPSPKLGW---MDEAMAIDPFNSDRLYGTATLYATNDLTWKDGS 415
 Db 573 tk-----iypggevehkkaqfkisvpggyvwdptn--dpsykgltsqleknkyaiaaydn 626
 QY 416 GOIHIAPMVKGLEETAVNDLISP-----PSGAP-----443
 Db 627 n-----lvwglepgaatstaptstptptptvtatptptptptptptptptptptpt 680
 QY 444 -----LISALGDL-----GGTHADVTAV-----P 463
 Db 681 vlyknnetsastgsirpfwkivngsssvdlrsvkirywtyvdgdkpqaavcdwaqlgas 740
 QY 464 STIFSPVFTTGS-VDYAELNPSIIVRAGSFPSSQPNDRHVAFTDGGKN-----514
 Db 741 nvtfnfvkissgvsagady-ylevgfsagagqlqpgktdgdiqvrfrknkdwsnynqadwds 799
 QY 515 WFQ-----GSEPGGVTTGGTVAASADGSFRVWAPGDP-----546
 Db 800 wlqsmtnygenakvtlylvgdvlwggqepgagatpaptstatptptatptptptptptptpt 859
 QY 547 -----GQPVVYAVGFGNSWAAASQGVPA-----NAQIRSDRVNPKTFYALSN-GT-- 589
 Db 860 vsatptpaptaspvg-gsywtbpsesygalkvwyangnlssptnlvlnpk--ikienyvtta 916
 QY 590 -----FYRSTDCGVN-----FOP-----VAAGLPSSGA-----VG 614
 Db 917 vdlrsrvkvrywytiddeatqsvsvassinpaiyidvfkvlganagadadyveigfksag 976
 QY 615 VMFHAVPGREGDILWLAASGLYHSTNGSSWSA-----ITGVSSAVNVGFGKSAPGS 666
 Db 977 vlaagstkeirlisickgsgsynqndysvrsansyienekvtgyiddvlvlgregrna 1036
 QY 667 STPAVFWVTIGGVTA-----YRSDCCGT-----WVLINDDQHOYGNWQAAYT 711
 Db 1037 qikvwyangnlsgmtavlnpkikienvgttavdlrsvkrywytiddeatqsvsvtsin 1096
 QY 712 GDHANLRRVYIGTNGRGI-VYGDIG 735
 Db 1097 payidvkvfkiganagadyveig 1121

RESULT 5
 AAP90615
 ID AAP90615 standard; protein; 745 AA.
 XX
 AC AAP90615;


```
QY 241 IPHKGVDPVNHVLIATSNLTGPGYDGSQGVWKFVSTGTTWTRISPVSPST----- 291
Db 302 qanvqafhnagikvymdv-----ynhtaeggtwtssqdtattatyswrgldn 348
QY 292 -----DTANDYF-GYSGLTIDRQHNTI--MVATQISWMPDTHIFRSTDDGATWTRW 341
Db 349 atyveltsnqfydyntlganfentvntvaqlivdsliaywntm---gvdgrf-----f 400
QY 342 DMTS-YPNRSRLRYLDISAEPMLTFCGQPNPPVSPKLGWMDAMAID----PFNSDRML 396
Db 401 diasvlgnsclngaytasa-----pocpngyynfdaadnsvainrillreftvrpaa 451
QY 397 YGTGATLYATNDLTWKDSGGQIH-IAPWVKGLEE-----TAVNDLISPPSAPLI 445
Db 452 gsgslldifa-----epwaignsyqlgfgpqgwswnlfrdslrqaqnel-----gsmti 502
QY 446 SALGDLGGTTHADVTAPVTITFTSPVFTTGTSDVYAEALNPISIIVR-----AGS 493
Db 503 yviqandfsgs-----snlfqsgsrspwnsinfidvhdgntlkdvyscngannsqawp 556
QY 494 FDPSSQPNDRHVAFTDGGK-----NWFQSEPGGVTTGTGV-----AASADGSRFVWAPGDP 546
Db 557 ygps-----dggststnyswdqgmsag---tgaavdgrraartgmafeml--sa 599
QY 547 GOPVYVAVGFGNSWAASQGVPAQAIRSDRVNPKTF-YALSNGTFFYRSTDDGVTFOFVAA 605
Db 600 gtpimg---gdeylrtlcnmnaaylndsawltyswttdqsnfytfagrliafrkabh 656
QY 606 GL-PSSGAGVGMFHAVPGKEGDLWLAASGLYHSTNG-----GSSSAIT--GVSSAVNVG 658
Db 657 alrpss-----wysgqltwyqpsgavadsnywnntsnayalayain-- 697
QY 659 FGKSAFGSSYPVAVVGTGTGGYTGAVRSDCGTTWVLINDDQHQYGNWGAITGDHANLR 718
Db 698 -gpslqdsnsiyayngwsssvfttlpappsgtqwyrvldt-----cdwndgastfvapgs 752
QY 719 RVYGTNGRGVYGDIG 735
Db 753 etlig--gagttgygcq 767

RESULT 8
AAU31850
ID AAU31850 standard; Protein; 2367 AA.
XX
AC AAU31850;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2341.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
```

```
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 519; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 2367 AA;

Query Match 3.7%; Score 150.5; DB 22; Length 2367;
Best Local Similarity 20.5%; Pred. No. 0.25;
Matches 160; Conservative 84; Mismatches 310; Indels 225; Gaps 43;

QY 3 TQP-----YTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWLPLLDWV 58
Db 1318 tqpnihdywnngigved-----wlaeerwlkfrfqtmdgmwensvswa 1360
QY 59 GWNWNGYGVVSIADPINTINKYAAVGMVTN--SWDPNDGAILRSSDQATQITPLPF 116
Db 1361 ddeneasigswgasdkagiirswavacdetsvksw-----agarae----- 1402
QY 117 KLGNNPGRGCMGERLAVDPNNDNILYFGAPSGKGLRSTD--SGATWSQMTNFDVGTYYIA 175
Db 1403 -----nvvgigtwarageqasgg--lwaggtsetgwagdkasggawtgaenqasggsa- 1455
QY 176 NPDTTGTGQSDIQGVVWVAFDKSSSL--COASKTIFGVGADPNPNPVFWSRGGATWQAV 233
Db 1456 -----lagna-igelwaagqasdgswpgqgqasgsvwgeaiggs--wt---gaenqas 1504
QY 234 PGAPTGFIPHKGVDPVNHVLIATSNLTGPGYDGS-----GDVWKF---SVTSGTWTRI 285
Db 1505 egswag--agagnmssvswagvvdqaggsagtsdqsgsgskprfedqasgegswa-- 1560
QY 286 SPVPSTDTANDYFYSGLTIDRQHPNTIMVATQISWMPDTHIFRSTDDGATWTRINDWTS 345
Db 1561 -----gagg-----qasggsmlgpe-----dqssgrswadtadqas 1591
QY 346 YPNRSLRYLDISAEPMLTFCGQPNPPVSPKLGWMDAMAIDPFNSDRMLYGTGATLYA 405
Db 1592 ggsr-lghvdqssggaw-----agtdqgs-----gggskprf 1622
QY 406 TNDLTK---W-----DSGGQIHIAPMVKGLEETAVN--DLISPPSGAPLISALGDL--GG 453
Db 1623 enqtteegswagaggqagggskvgpedqssgrswansgdqi---sgglvlgvdqanggs 1679
QY 454 FTHADVTAVPSTTFTSPVFTTGTG--VDYAEALNPSSI--VRAGSFPDSSQPNDRHVAFTS 509
Db 1680 wtga---ghpasvpgkpfiedqvsgrgswadareqvvgdsrlglrdqssgds---wagtg 1733
QY 510 DGCKNWF---QSEPGGVTTGTTVAASADGSRFVWAPG-----DPGQPVVYAVG 555
Db 1734 dqasgwfcvcpqsgtnggswggasqdvvggr-----pgptnqssagdsdpsgsv----- 1784
QY 556 FGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFFY-----RSTDGQVT--FQPVAAGLPS 609
Db 1785 sgscwtgagavdqagcscpggfedqai-----gggfwpgagdqgtgggrpsgedqssgigs 1840
```


Db 279 glavqngsv-----ctngtnalvspilpgsvltisgnvdasgksctniissggqvaiaqdgqav 334
QY 322 WPTIIIFRSTGDGATWRIWMTSYPNRSLRYVLDISAEFWLTFGVQPNPVPSPKLGWM 381
Db 335 qgatvsgnggggt-----vrigdfigq-llc-----lpnasqtli 370
QY 382 DEAMADFPNSDRMLYGTGATLYA-TNDLTKWDS-----GGQIHAPWVKGLEET- 430
Db 371 dsnsvv---kadalltngngtviwaddstrfsgnisagggmtggn-----ggfvets 420
QY 431 -AVNDLISPPSGAPLISALCDLGCF-----THADVTAVPSTIFTSVPVTTGTSVD 479
Db 421 gakslnvddtarvntfatmgelgtwllpleiivgttdlldapkv---svltittsld 477
QY 480 ----YAEINPSIIVRAG-SPDPSQPN-----DRHVAFTDGGKNNFQSGSEPGGVTT 526
Db 478 ngnvilqadqsiavqanfsadpsapgnltfdspitidalifslgtgsiifantgp--int 535
QY 527 GGTVAAS-----ADGSRFVWAPGDCQPVVYAVGFGNSWAASQGVPAQAQI-- 572
Db 536 gntlvtsfpntlnfdnkignanttftagp-----ydiyfrksvnggfdllignanfvy 588
QY 573 ---RSDRVNPKTYALNSNGFYRSTD---GGVTFOFVAAGLP---SSGAVGVMEHAVP 621
Db 589 fddgagittlpksgfvtateiyvgnndivtgnqifdgvfylqpvnltsagsviftnni 648
QY 622 GKEGDLWLAASSGLYHSTNGSSWSAITGVSSAV-----NVGFCK-----SAPGS 666
Db 649 llngsiqvtaqviv--sqpsaslsavelasdvllnagqnvsgnintrgnvdiqalgn 706
QY 667 SYPAVFVGTIGGVGAYRSDDCGTTWLVINDQHQYGNWQQAITGDHANLRVYIGTN- 725
Db 707 istgsivtsfpfgnagnvlnnaggt--littgyietsgtnggdvttssgntstayidtrg 764
QY 726 -GRGIVYGDIGGAPS 739
Db 765 fgdgleidsiggavs 779

RESULT 13
ID ABB59423
XX ABB59423 standard; Protein; 1472 AA.
AC ABB59423;
XX ABB59423;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 5061.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
XX WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL03526.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX Disclosure; SEQ ID NO 5061; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1472 AA;
SQ

Query Match 3.6%; Score 144.5; DB 22; Length 1472;
Best Local Similarity 21.2%; Pred. NO. 0.34;
Matches 191; Conservative 91; Mismatches 342; Indels 275; Gaps 45;

QY 14 GGGGFVDGIVFNEG-APGILYVRT-----DIGMYRWDAANGRW-----IP 53
Db 237 ggggrgggfvrsrgggrrmggrtggprgrgsgggpgay----gsrggnanedhheve 292
QY 54 LLD-----WVGNNMGYGVVS-----TAADPINT 78
Db 293 lwdntiaaekqqaahddawgdnneeeyegskdkvfttsnlatgsaanvsgtasv 352
QY 79 NKVAAVGMVNTSWDP-----NDGAILRSSDQATWQITPLPFLKGGNPGMGGERL 131
Db 353 tavpaagaetisa-ppglehqlvqqgshleesssgpaavtp-patlsagattplldysa 410
QY 132 AVDPNNDNILYFCAPSGKGLWRSTD--GATWSQMTNE---PDVGTVIANTDITGQSD 186
Db 411 avsnpppqiqsgtqsgagtgsaaagggagstpsfvsasptdfssaasaaatlvhqag 470
QY 187 IQGVVWVAFD--KSSSLGQASKTIFGVADPNPNPVFWSRGGATWQAVP--GAPTGFIP 242
Db 471 kqqqlqqqttpikpsatlsveqsyfnsasq-----gvspsgsvpqsapagyaq 520
QY 243 HKGVFDPVNHVLYIATISNTGG--PYDSSGDVWKFVTSGTWT-----RI-- 285
Db 521 ----npv--aaysqtstsvgsqypntyavfasgtaagtaegsqgqppqirrarvkl 573
QY 286 ---SPVPST-----DTANDYFGY-----SGLTIDRQHPNT 312
Db 574 ppskipasavempgdnalninnigylvqfgaldfgtdgdgfeipekvsgsfidgq---- 629
QY 313 IMVATQISWMPDTIIFRSTDGGATWRIWMTSYPNRSLRYVLDISAEFWLTFGVQPNPP 372
Db 630 ----qqqqqpd-----dygksqg-----qqqvtlaaglaq----- 655
QY 373 VPSPKLGMDEMAAIDFPNSDRMLYGTGATLYATND--LTKWD-----SGQIHIAPMV 424
Db 656 --ssqtsdalaanaagytststsqqqgqgvsavnatidqltksdpyggtgsgsnaqyqs 713
QY 425 KGLEETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSTPVTFTGTSVDYAEIN 484
Db 714 sgaskta-----sgfpttap-----ggysstyanvqssvans-----yqqqq 751
QY 485 PSIIVRAGSFPDPSQPN-----DRHVAFTDGGKNNFQSGSEPGGVTTGTTVAASADGSR 538
Db 752 -----ygsyqpsyyqqagsgagsgtgvsggggtatqnpvgsssqnsgnassay 805
QY 539 FVWAPGDP-----GQPVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTYALNSNGFYR 592
Db 806 ltsgystpqsayqsgsvyngtvlslns-sgfsagsnas--sgyanfssaaklkdatas 862
QY 593 ST---DGVVTFQPVAAAGLPSSGAVGVMEHAVPVGKGDWLWLAASSGLYHSTNGGSSWAIT 649

XX
AC AAR47578;
XX
DT 19-JUL-1994 (first entry)
XX
DE Flocculation protein of *Saccharomyces cerevisiae*.
XX
DE Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*;
KW flocculation protein; enzymatic process; fermentation;
KW biodegradation; catalysis.
XX
OS *Saccharomyces cerevisiae*.
XX
PN W09401567-A.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-EP01763.
XX
PR 08-JUL-1992; 92EP-0202080.
PR 14-DEC-1992; 92EP-0203899.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
PI Klis FM, Schreuder MP, Toschka H, Verrips CT;
XX
DR WPI: 1994-035071/04.
DR N-PSDB; AAQ54029.
XX
PT Immobilisation of enzymes to microbial cell wall - by prodn. of
PT fusion protein of enzyme linked to anchoring protein
XX
PS Example 10; Page 59-64; 99pp; English.
XX
CC The flocculation protein is used in a method to immobilise enzymes
CC to a microbial cell wall. The coding sequence (FLO 1) is used in
CC the production of a recombinant polynucleotide which comprises a
CC structural gene encoding a protein with catalytic activity and at
CC least part of a gene encoding at least the C-terminus of a protein
CC capable of anchoring in a eukaryotic or prokaryotic cell wall. The
CC anchoring fragment or protein is selected from alpha agglutinin,
CC AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC proteinase of lactic acid bacteria. The recombinant polynucleotide
CC preferably also comprises a sequence encoding a signal peptide to
CC ensure secretion of the expressed product. The signal peptide is
CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.
CC The host microorganism can be used for performing enzymatic
CC processes on an industrial scale.
XX
SQ Sequence 894 AA;

Query Match 3.5%; Score 143; DB 15; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.22;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGAPGILYVTRDGGMYRWDAANGRWIPLLDWVG----- 59
Db 53 tlynaaymaygya-----sktlgsvggtdtsidyn-----ipcvsssgtfpcpqe 99
QY 60 --WNNNGYNGVWSIAADPINTKNK--WAA--VGMYTNSTD----- 93
Db 100 dsgynwgckgmgaacs-----nsqgiaywstdlfgfytbtntvltmgtgyflbpqtsytfk 155
QY 94 ---PNDGAILRSSDQGATW-----QITPLPFKLGGMFGRGMGERLAVDPNNDNL 141
Db 156 fatvddsaill--svggatafnccaqppitstnftidgikpwgg-----slppniegtv 208

QY 142 YFGA----PSGKGLMRSTDGATWSQMTNFPDVGTYIANTPTDTFTGYOSDIQGVVWAFDX 197
Db 209 ymyagyyyym-kvysnavswgtlplsvtlpd-gttvsd-----dfegvyv-sfd- 255
QY 198 SSSSLGQASKTIFVGADPNPNVFNWSDRGATWQAVPGAPTGFIPHKGVDPDPVNHVLYIA 257
Db 256 --ddlsqsnct---vpdpn-----yavsttttpeptwtgtfststemttv 297
QY 258 TSNLTGGPYD-----GSSGDVMRFSVTSGTWTRISPVSTDTANDYFGYGLTID-- 306
Db 298 tgnvgvptdetvivrptsegli-----sttpevptgtfststevit-igtngqptdet 353
QY 307 ----ROHPNTIMVATOISWMPDPIIFRSDGGATWTRIMDWTSYPNKRSRYVL----- 355
Db 354 vivirptseglisttpevptgtfstststse-----mtvtgtnqgptdetvivrptseg 409
QY 356 --DISAEPWL-TFVGQPNPVPSPKLGWMDENAIIDPFNSDRMLYGTGATLYATNDLTWK 412
Db 410 lvtttteptwtgtf-----tststemstvtgtnqplptdetv-- 444
QY 413 DSGGQIHIAIPMVKGLEETAVNDLISPPSGA-----PLISALGDLGGFTHADVTAV 462
Db 445 -----ivvktpttaisslsssssgqitssitssrpiitpfypsnq-tsvissv 493
QY 463 PSTIFTSPVTTGTSDVYAEELNPSIIVRAGSFPDPSSQPNDRHVAFTDGGKNNFQSGSEP 522
Db 494 issvstssiftspvissvisssttststisfesskssvptssstsgsse-setssag 552
QY 523 GVTGTGTVAAASADGSRFVWAPCDGQPVVYVAVGFGNSWAASQGVANAOIRSDRVNPKTF 582
Db 553 svsssssfissessksp---tyssslplvtststgetassl-ppatttktseqtlvt- 607
QY 583 YALSNGTFFRSTDGGVTFQFVAAAGLPSGAVGVNMFHAFVG--KEGDLWLAASS----- 633
Db 608 -----vtsceshvctesispavststvtv-----tvsgvtteytwcpisttettkqt 654
QY 634 -GLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPAVF--VVGITGGVTAIRSDDCG 690
Db 655 kgteqtdttktktgtvtvtvttisscesdvcsktas----pavststatingvttey----- 704
QY 691 TTWVLIN 697
Db 705 ttwcpis 711

Search completed: July 2, 2002, 09:15:50
Job time: 238 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:46 ; Search time 39.87 Seconds
(without alignments)
453.347 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPYTWSNVAIGGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	162.5	4.0	772	2	US-08-410-784A-5
2	158.5	3.9	774	4	US-09-346-237-8
3	152.5	3.8	1751	4	US-09-136-574A-44
4	151.5	3.8	776	4	US-09-346-237-4
5	148.5	3.7	750	6	5457037-3
6	148.5	3.7	751	6	5457037-5
7	148.5	3.7	776	4	US-09-346-237-7
8	143	3.5	894	3	US-08-362-525-22
9	143	3.5	894	3	US-08-971-692-15
10	141	3.5	720	4	US-09-296-284-25
11	141	3.5	754	4	US-09-296-284-4
12	140.5	3.5	1000	4	US-09-352-159-25
13	140.5	3.5	1000	4	US-09-352-168-25
14	136	3.4	2628	2	US-08-570-311-14
15	135.5	3.4	1205	4	US-09-352-159-29
16	135.5	3.4	1205	4	US-09-352-168-29
17	132.5	3.3	626	6	5268290-2
18	132.5	3.3	866	1	US-08-386-727-8
19	132.5	3.3	866	2	US-08-600-452A-8
20	132.5	3.3	1395	4	US-09-540-245A-15
21	131.5	3.3	1026	2	US-08-542-003-6
22	131.5	3.3	1026	2	US-08-322-760A-6
23	130.5	3.2	862	1	US-08-325-267A-4
24	130.5	3.2	1326	4	US-09-147-236-5
25	130	3.2	1026	1	US-08-194-290-7
26	128	3.2	1042	3	US-08-928-361B-11
27	128	3.2	1837	3	US-08-928-361B-5

28	125.5	3.1	538	3	US-09-040-005-2	Sequence 2, Appl
29	125.5	3.1	538	4	US-09-522-217-115	Sequence 115, App
30	125.5	3.1	551	2	US-09-033-537A-1	Sequence 1, Appl
31	125.5	3.1	565	4	US-09-142-623-11	Sequence 11, Appl
32	125	3.1	735	2	US-08-313-185-48	Sequence 48, Appl
33	125	3.1	735	2	US-08-459-499-9	Sequence 9, Appl
34	125	3.1	735	2	US-08-459-499-12	Sequence 12, Appl
35	125	3.1	735	3	US-09-082-614A-48	Sequence 48, Appl
36	124	3.1	553	2	US-08-661-052-16	Sequence 16, Appl
37	124	3.1	553	4	US-09-188-082-16	Sequence 16, Appl
38	124	3.1	746	1	US-08-476-519-11	Sequence 11, Appl
39	124	3.1	746	5	PCT-US95-09323-11	Sequence 11, Appl
40	124	3.1	777	1	US-08-476-519-2	Sequence 2, Appl
41	124	3.1	777	5	PCT-US95-09323-2	Sequence 2, Appl
42	124	3.1	3031	1	US-07-689-008-2	Sequence 2, Appl
43	123	3.0	1160	3	US-08-808-599A-24	Sequence 24, Appl
44	122.5	3.0	802	4	US-09-147-236-4	Sequence 4, Appl
45	122.5	3.0	3519	4	US-09-428-517-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-410-784A-5
; Sequence 5, Application US/08410784A
; Patent No. 5912413
; GENERAL INFORMATION:
; APPLICANT: MYERS, ALAN M.
; APPLICANT: JAMES, MARTHA G.
; TITLE OF INVENTION: ISOLATION OF SU1, A STARCH DEBRANCHING
; TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
; TITLE OF INVENTION: SUGARY 1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,784A
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heine, Ph.D., Holliday C.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: ISU-002XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-2290
; TELEFAX: 617-451-0313
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
US-08-410-784A-5

Query Match 4.0%; Score 162.5; DB 2; Length 772;
Best Local Similarity 19.8%; Pred. No. 0.00071;
Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;

QY 51 WIPLLDWGNNMGYN---GVWS---IAADPINTNKVAAVGMVMTNSWDPNDGAILRSSD 104
DB 108 WGP--NWPYASNGKSGACVSDVDANGDREPNKLLLDPPYAQESQDP-----155
QY 105 QGATWQITPLPFLKGNMPCRGMERLAVDPNNDNLIYFCAPSGKGLWRSTDGATWSQM 164
DB 156 -----LNPSNQ-----GNVFASAHRTTDSGI-----178
QY 165 TNPEDVGTIANPTDTTG-----YQSDIOGVVWVAFDKSSSLG-----203
DB 179 --YAPKGVVLVPSTQSTGKTPTRAQKDDVIYEVHVG-----FTEQDTSIPAQYRGTYG 231
QY 204 ---QASKTIFVGVA-----DPNNPV-----FWSRDGGATWQA 232
DB 232 AGLKASYLASLGTAVEFLPVQETONDANDVVPNSDANONWGYMTENYFSDRRYAYNK 291
QY 233 VPGAPG-FIPKHGVEDPVNHLVLIATSTGGPYDSSGDYWKFSVGTWTRISVPST 291
DB 292 AAGGPTAEFOAMVQAFHNAGIKYMDV-----VYNHTAEGGWTSSDPTAT 338
QY 292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATOISWMPDTIIFRSTD 332
DB 339 IYWRGLDNATYIELTSGNOYFYDNTGIGANFNTYVTAQNLIVDSVAYWNTM---GVD 395
QY 333 GGATWTRIMWTSYPNRSRLRYLIDSAEPWLTFGVQPNPPVPKLGWMDMAIDPFNS 392
DB 396 GFR-----FDLASVLGNSCLNAVHASA-----PNCNPGGYNFDAAASNVAI-----436
QY 393 DRMLY-----GTGATLY-----ATNDLTW-DSGQIHIAPMYKGLEETAVNDL 435
DB 437 NRILRETVRPAAGGTWICLRNLGPSAATR--TSWVDSRRVVRVEMSVPRQLRQONEL 494
QY 436 ISPPSGAPLISALDGGFTHADVTAVPSTIFTSPVFTTGSVDYABLNPISIIVR-----490
DB 495 ---GSMTIYVTDANDFSGS-----SNLFOSGSRSPWNSINFIDVHDGMLKDVYSC 543
QY 491 --AGSDPPSQPNDRHVAFTDGGKWFQSEGGVTTGGTV---AASADGRF-VWAPG 544
DB 544 NGANNSQASYPGSDG-----GTSTNYSWDQMSAG---TGAADVQRRAAARTGMAFEMLSAG 596
QY 545 DP-----GOPYVYAVGFGN-----SWAASQG---VPANAQIRSDR---VNPX 590
DB 597 TPLMOGGDEYLRLOCNNNAYNLDSSANWLTYSWTTDQSNFYTFQAQLIRSAHIPLRPS 656
QY 581 TFYALSNGTFTYRSTDGGVTQFPVAAGLPSSGAGVGMFHAVPGKGDLLWLAASSGLYHSTN 640
DB 657 SWYSGSLTWY-----QP-----SGAV-----ADSNWNNTS 683
QY 641 GGSWSAITGVSSAVNVGFKSAPGSYPVAVFV---GTIGGVTCAYESDDCGTTWLIND 698
DB 684 NYAIAVAINGPSL-----GDSNDSIYVAYNGWSSSVTTLPPAPSGTQWYRVD 732
QY 699 DQHQYGNWGOAITGDHANLRVRYIGYNGRIGVYGDIG 735
DB 733 T-----CDWNDGASTFVAPGSETLIG--GAGTTYGQCG 763

RESULT 2
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629,200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A

; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Favobacterium odoratum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(774)
; OTHER INFORMATION: Isoamylase
US-09-346-237-8

Query Match 3.9%; Score 158.5; DB 4; Length 774;
Best Local Similarity 18.8%; Pred. No. 0.0014;
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;

QY 76 INTNKVWAA-----VGMVMTNSWDPNDGAILRSSDQATWQITPL 114
DB 35 INPNKLGAAAYDATKANVTFKVYSKATRIELYLSTATGSAEKAKYVMTNSGGIWSVT-I 93
QY 115 PFKLGCMPCRGMERLAVDPNNDNLIYFCAPSGKGLWRSTDGATWSQMTNFPDVGTYI 174
DB 94 P---TSTLSQGLG-----GTYLYGYRANGPNW---PYNASTKGSSL-----130
QY 175 ANPTDTGTQSDIOG-----VWVAFDKSSSLGQASKTIFVGVADPNNP 219
DB 131 -----GFTSDVDAAGNRENPNKLLSDPYALELSHDPATTATM-----TNGS 170
QY 220 VFWSRDGGATWQAV---PGAPTGF-----PHKGVFDPV---NHLVLIATSTNG 262
DB 171 IYAS---GATYRINDSGSSAPKGIIVLAGDTQATGTPTRAKDDVVYEAHVHGLTWMNDS 227
QY 263 -----GPYDSSGDYWKFSVTSCTWTRISPV-----PSTDANDYFYGSLTI 305
DB 228 ITAARYRTYKAGLKAALAAALGVTAIEFLPVQETQNTDNDPSTSGDNYNGYMTLNY 287
QY 306 ---DRQ-----HPNTIMVATQISWMPDTIIFRSTDGATWTRIW 341
DB 288 FAPDRRYAYDKTPGPTREFKENVKAFHDNGIKVLVD---VYVNHHTGEGGASPT- 339
QY 342 DWTSYPNRSLR-----YVLDISAE--PWLTEGVQPNPPVPS-----PKLGWMDAM 385
DB 340 DKTYNITSFRLDNTYYSLSLTADFQNSWNTGVGNYNTRNTIAQNLIVDSLAYWRDKL 399
QY 386 AIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHIAPMYKGLEETAVNDLIS--PP----439
DB 400 GVDGYRFDLASVLGNSCQHCFCNFKMDAG-----NALNRIVAELEPPRPAT 445
QY 440 ---SGAPLIS---ALG---DLGGF-----THADYVAVPSTIFTSPVFTTGSVDYA 481
DB 446 GSGVDLIIAEPWAPGNSYQVGGFPGSWEWNGAYRDVVRQAQONKLGSAVITG-----499
QY 482 ELNPSIIVR--AGSFDPSQPNDRHVAFTDGGKWNFGQ---SEPGGVVTTGGTVAASADGS 537
DB 500 -----QMATRFAGSSD-----LYGDDGRKPWHSVNFITAHDGFTLKDLYSCSNKN 545
QY 538 RFVW---APGDPQPVVYAVGFGNSW-----AASQ-----GVV-----ANQA 571
DB 546 NQWMPYGPSDGE-----DNNNSWDGGIADQDKAARNGMALMLLSAGVPMIVGDEA 599
QY 572 IRS-----DRVNPKTFY-----ALSNGTYRSTD-G 596
DB 600 LRSMCNPNPNYLDSSANWLNWSRTDQNNFQSFKAMIAFRKAHPALRPANFYSSVDNN 659
QY 597 GVTFPQVAAGLPSSGAGVGMF-----HAYPGK--EGDLWLAASSGLYHSTNGSSWSAIT 649
DB 660 GNYMEQLRWFKPDGGVADATYFNDANNHATAWRIDGSEFGDTSATYVAHN---ANSA-- 714

QY 650 GYSSAVNGFGKSA^PGSSYP^AVFVVG^TIGGVT^GAYRSDDC^TTWVLINDDQH^OYGNGW^G 708
||| ||| | | | | : : : ||| :
Db 715 -----QVNFTLPNPGAGKSWRYTDTCGWAEGASQVAPGSE-ALVGGENTAYGLCGR 766

RESULT 3
 US-09-136-574A-44
 ; Sequence 44, Application US/09136574A
 ; Patent No. 6294366
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Farrington, Graham K.
 ;
 ; Anderson, Palge
 ; Gibbs, Moreland
 ; Bergquist, Peter
 ; Daniels, Roy
 ; Morgan, Hugh W.
 ; Williams, Diane P.
 ;
 ; TITLE OF INVENTION: Compositions and Methods for
 ; Treating Cellulose Containing Fabrics Using Truncated
 ; Cellulase Enzyme Compositions
 ;

CELLULOSE ENZYME COMPOSITIONS

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

```
, ,  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/O9/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
,
```

PRIOR APPLICATION DATA: US 08/932,571
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

```

: INFORMATION FOR SEQ ID NO: 44:
: SEQUENCE CHARACTERISTICS:
:     LENGTH: 1751 amino acids
:     TYPE: amino acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:     MOLECULE TYPE: protein
:     SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

```

Query Match	3.8%	Score 152.5;	DB 4;	Length 1751;
Best Local Similarity	18.7%	Pred. No. 0.012;		
Matches 184;	Conservative	101;	Mismatches 349;	Indels 351;
				Gaps 49;

[illegible]

344	Db	AAFLACVADWSSCCSNKTKYLNFAKSQIDYALGST--GRSFVVVGFTNTYPOPHHRN--	400
141	Qy	LYFGAPSGKGLMRSTDGATWQMTNFPD-----VGTIIANTDPTTGYQSDIOGVVW--	192
401	Db	-----AHSSWANSMKIPEYHRHILYGALVGGPGSDSDNDIIDYYQNE	444
193	Qy	VAFDKSSSLGQASKITIFGVGADPNNPFWSRDGGATWQAVPGAPTGFIPHKGVDPVNH	252
445	Db	VACDYNAGIVGALAK-----MYLYGGEPID--DFKAIETPTND	481
253	Qy	VLXIAT--SNTGGPYDSSGDYWKFSVTSGTWTTRISPVSPDTAN-DYF-----GY	300
482	Db	EIFVESKFGNSQP--NYTEVISYIYNTGW---PPRVTDKLSFKFYIDLTELIGAQY	534
301	Qy	SGLTIDRQHPNTIWTATQISWMPDITIIFRSTDGGATWTRINDWTSYPNRSLRYVL-DISA	356
535	Db	S-----PDVVKVDT-----YYIEGGKISGPYVMD---KRNRIYYVLVDFSG	572
360	Qy	EPWLTFGQVQNPVPSPKLGW---MDEAMATDPNPSRMLYGTGATLYATNDLTWKDWSG	415
573	Db	TK-----IYPGGEVHEHKQAQPKISVPGQYPNWDPIN-DPSYKGLTSQLEKKNYTAA	626
416	Qy	GQTHIAPWKVGLBETAVNDLISP-----PSGAP-----	443
627	Db	N-----LVWGLEPGAATSTPAPTSTPTPTPTPTVTATPTPTPTPTGSPGTSGVK	680
444	Qy	-----LISALGDL-----GGTHADTVAV-----P	463
681	Db	VLKNNETSASTGSRPWFKIVNGSSSVDLRVRKIWTYVDGDKPOSACVDWAOIGAS	740
464	Qy	STIFTSVFVTGTGS-YDYAELNPSTIVRAGSPDPSQPNRHFVAFSTDGKN-----	514
741	Db	NVTFNFKLSSGVSQADY-YLEVGFSGAGQLQPKDGTGDIQVRENKNDWSNIOQADWS	799
515	Qy	WFO-----GSEPGGVITGGTVAASADGSRFVWAPGD-----	546
800	Db	WLOSMTNYGENAKVTIYVQVLVWQEPGGGAPTATPTPTPTPTPTPTPTPTPTPTPT	859
547	Qy	-----GQPVVYAVFGNSWAASQVPA-----NAQIRSDRNPXTFFVALSN-GT--	589
860	Db	VSATPTPAPTASPVG-GSVWTPSESYGALKVWYANGNLSPTNVLNPK--IKIENVGTTA	916
590	Qy	-----FYRSTDGGVT-----FQP-----VAAGLPSSGA-----VG	614
917	Db	VDLRVRKRVWYTIDGEATQSVSVASSINPAVIDKVKLGANAGADYYVEIGFKSGAG	976
615	Qy	VMPEHVPKGEDLWLAASSGLYHSTNGSSWSA-----ITGVSSAANNVCFKSPAGS	666
977	Db	VLAAGOSTKEIRLSIOKSGSNGSQNDYSVRANSYIENEKVITGYIDDLVWVGRPGRNA	1036
667	Qy	SYPAFVWVGTIGCVTGA---YRSDCCGT-----WVLINDDQHQYGNWGOAIT	711
1037	Db	QIKVWYANGNLGSMTVNLNPKIKIENVGTTAVDLRVRKRVWYTIDGEATQSVSYTSSIN	1096
712	Qy	GDHANLRVYIGNNGRI-YVGDIG	735
1097	Db	PAYIDKVFVKLGANAGADYYVEIG	1121

```

RESULT      4
US-09-346-237-4
; Sequence 4, Application US/09346237A
; Patent No., 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346, 237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA. 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353

```


Db 631 ALRPSS-----WYSGSLTWYQPSGAVADSNYWNNTSYAIAIYAIN-- 671
QY 659 FGKSAFGSSYPVAVVVGTTGGVTGAYRSDCGTTWVLINDQHOYGNWGAQITGDHANLR 718
Db 672 -GPSLGDNSIYAYNGWSSVFTFLPAPPSTGTWYRVTD-----CDWNGDASTFVAPGS 726
QY 719 RYVIGTNGRGIYVDIG 735
Db 727 ETLIG--GAGTTYGCG 741

RESULT 6
5457037-5
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1.797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO: 5:
; LENGTH: 751
5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;
Best Local Similarity 20.6%; Pred. No. 0.0072;
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRDAANGRWIPLLDWYGNWNGYGVVSIADPINTNKVWAAVG 86
Db 71 GITGAVY-----YGYRANGPNWPIYASNGWKGSGAGFSDVDANGDRFNPNKLLDPY 122
QY 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGGNPNRGMGRERLAVDPNNDNILEFGAP 146
Db 123 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157
QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 158 KGVVLVPSTQSTGKTPTRAQKDDVIEYVHVRGFTQDTSIPAQYRGTYTGAGLKASYLAS 217
QY 202 LGOASKTIFGVAD-----PN---NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 218 LG-VTAVEELPVOETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGTAEF 276
QY 241 IPHKGVDFPNVHVLXIATSNTPGPDGSGDVWKFVSTGCTWTRISPVST----- 291
Db 277 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTIYISWRGLDN 323
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATQISWNPDTIIFRSTDCGATWTRIW 341
Db 324 TTYIELTSGNQFYDNTGIGANFTYNTVAQNLI VDSLAYWANTM---GVDGFR-----F 375
QY 342 DWTS-YPNLSRLVLDISAEPLTFGVQPNPPVPSPLGWDEMAID-----PFNSDRML 396
Db 376 DLASVILNGLNGAYTASA-----ENCNPGGYNFDAAASNVAINRILREFTVRPAA 426
QY 397 YGTGATLYATNDLTWKDSSGOI-H-IAPMVKGLEE-----TAVNDLISPPSGAPLI 445
Db 427 GSGGLDLFA-----EPAWIGNSYQIGGFGQGWSEWNGLFDRSLRQONEL-----GSWTI 477
QY 446 SALGDLGGFTHADVATVPSTIFTSVFTTGTSDYAEALNPSIIVR-----AGS 493
Db 478 YVTQDANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVSCYNGANNQAMP 531
QY 494 FDPSSQPNDRHVAFSTDGK-----NWFQSGEPGGVTGTGTV---AASADGRFVWAPGDP 546

Db 532 YGPS-----DGGTSTNYSHWQMSAG---TGAVDORRAARTGMFAEML--SA 574
QY 547 GQPVVYAVGFGNSWAASQGVPANAQIRSDRVNPKTF-YALSNGTFYRSTGGVTFQPVAA 605
Db 575 GTPLMQG---GDEYLRTLQCNNAYNLDSSANWLTYSWTTDQSNFYTFQRLIAFRKAHP 631
QY 606 GL-PSSGAVGVMFHAVPGKEGLWLAASGLYHSTNG-----GSSWSAIT--GVSSAVNVG 658
Db 632 ALRPSS-----WYSGSLTWYQPSGAVADSNYWNNTSYAIAIYAIN-- 672
QY 659 FGKSAFGSSYPVAVVVGTTGGVTGAYRSDCGTTWVLINDQHOYGNWGAQITGDHANLR 718
Db 673 -GPSLGDNSIYAYNGWSSVFTFLPAPPSTGTWYRVTD-----CDWNGDASTFVAPGS 727
QY 719 RYVIGTNGRGIYVDIG 735
Db 728 ETLIG--GAGTTYGCG 742

RESULT 7
US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; EARLIER FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

Query Match 3.7%; Score 148.5; DB 4; Length 776;
Best Local Similarity 20.6%; Pred. No. 0.0075;
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps 42;

QY 27 GAPGILYVRTDIGMYRDAANGRWIPLLDWYGNWNGYGVVSIADPINTNKVWAAVG 86
Db 96 GITGAVY-----YGYRANGPNWPIYASNGWKGSGAGFSDVDANGDRFNPNKLLDPY 147
QY 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGGNPNRGMGRERLAVDPNNDNILEFGAP 146
Db 148 AOEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201
Db 183 KGVVLVPSTQSTGKTPTRAQKDDVIEYVHVRGFTQDTSIPAQYRGTYTGAGLKASYLAS 242
QY 202 LGOASKTIFGVAD-----PN---NPVFW-----SRDGGATWQAVPGAPTG-F 240
Db 243 LG-VTAVEELPVOETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGTAEF 301
QY 241 IPHKGVDFPNVHVLXIATSNTPGPDGSGDVWKFVSTGCTWTRISPVST----- 291
Db 302 QAMVQAFHNAKIKYMDV-----VYNHTAEGGTWTSDDPTTIYISWRGLDN 348
QY 292 -----DTANDYF-GYSGLTIDRQHNTI---MVATQISWNPDTIIFRSTDCGATWTRIW 341


```
Db 349 TTYELTSGNYFYDNTGIGANFNTYNTVAQNLIIVDSLAWNTM---GVDGFR-----F 400
QY 342 DWTS-YNRSRLRYVLDISAEPWLTFTGQPNPVPSPKLGWDEAMID----PNSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCNGGYNFDAAASNVALRILREFTVRPAA 451
QY 397 YGTGATLYATNDLTWKDMSGQI-H-IAPMWKGLBE-----TAVNDLISPPGAPLI 445
Db 452 GSGCLDLFA---EPMAIGNSYQLGFGPQGWSEWNGLFPRDSLRLQAQNEL-----GSMTI 502
QY 446 SALGDLGGFTHADVTAVPSPFTFTGTSVDYAEALNFSIIVR-----AGS 493
Db 503 YVTQDANDFSGS---SNLFSSGSRSPWNSINFIDVHGMTLKDYISCGANNQAWP 556
QY 494 FDPSSQPNDRHVAFTSDGGR---NMFQSEPGVGTGGTV---AASADGSRFVWAPGDP 546
Db 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVORRARTGMAFEML---SA 599
QY 547 GQPVVAVGFGNSWAASQGVYANQIRSDRVNPKTF-YALSNCTFYRSTDDGGVTFQPVAA 605
Db 600 GTPLMQG---GDEVLRTLQCNNAYNLDSSANWLTYSWTTDQSNFYTFQORLIIAFRAHP 656
QY 606 GL-PSSCAVGMVHAPVCKEGDLWLAASSGLYHSTNG-----GSSWSAIT--GYSSAVNNG 658
Db 657 ALRPS-----WYSGSLTWYQPSGAVADSNYWNNTSNYAIAYAIN-- 697
QY 659 FGKSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHQYGNWQAITGDHANLR 718
Db 698 -GPSLGDNSIYVAYNGWSSSVFTLPAPSPGQWRYVTD---CDWNGDCASTFVAPGS 752
QY 719 RVYIGNRGIVYGDIG 735
Db 753 ETLIG--GAGTTYGQC 767

RESULT 8
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
```

```
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-525-22

Query Match 3.5%; Score 143; DB 3; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.023;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG----- 59
Db 53 TYSNAAYMAYGA-----SKTLGSGVGGQTDISIDYN-----IPCVSSTGTFPCPQE 99
QY 60 --WNNNGYGVVSIADPINTNKV--WAA--VGMVTNSWD----- 93
Db 100 DSYGNNGCKGMGACS---NSQIAVWSTDLFGFYTPTNVTLEMTCYFLPQTGSGYFK 155
QY 94 ---PNDGAILRSDQATW-----QITPLPFKLGNNPGRGMRGELAVDPNNDNIL 141
Db 156 FATVDDSAIL--SVGGATAFNCCAAQOPPTSTNFTDGIKPMGG-----SLPPNIEGTV 208
QY 142 YFGA---PSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDITGYQSDIQGVVWAFDK 197
Db 209 YMTAGIYYPM-KVVISNAVSWGLPIPSVLPD-GTVSD-----DEGVYI-SFD- 255
QY 198 SSSSLGQASKTIEFVGVADPNPNVFWSRDGGATQWQVPGAPTGFPHKGVDPDVNHLVYA 257
Db 256 --DDLQSQNT---VPDPN-----YAVSTTTTTEPTWTGTSTSTEMTV 297
QY 258 TSNITGGPYD-----GSSGDVWKFESVTSQWTRISVPVSTDTANDYFGYSGLIID-- 306
Db 298 TGTNGVPTDETIVIRTPTEGLI---STTEPTGTFTSTSTEVTT-ITGTNGQPTDET 353
QY 307 ---RQHPNTIMVATQISWMPDIIIFRSTDGGATWTRINDWTSYPNRSLRYVL----- 355
Db 354 VIVIRPTSEGLISTTTEPTWTGTFTSTSTB---MTVTGTNGQPTDETIVIRTPTESEG 409
QY 356 --DISAEPWL-TFGVQPNPPVSPKLGWMDMAIDAIDPFNSDRMLYGTGATLYATNDLTKW 412
Db 410 LVTTTTEPTWTGT-----TSTSTEMSTVTGTNGLPTDETIV-- 444
QY 413 DSGGQTHIAPMWKGLEETAVNDLISPPGA-----PLISALGDLGGFTHADVTAV 462
Db 445 -----IVVKTPTTAISSLSSSSQITSSITSSRPITTFYPSNG-TSVISSV 493
QY 463 PSTIFTSPVETTGTSVDYAEALNPSIIIVRAGSPDPSPQPNDRHVAFTSDGKKNWFQSEPG 522
Db 494 ISSSVTSSLIFTSSPVISSSVISSTTSTTSIFSESSKSSVPTSSSTSSSE-SETSSAG 552
QY 523 GVTGTGTVAAADGSRFVWAPGDPGVVYAVGVFGNSWAASQGVVAPNAQIRSDRVNPKTF 582
Db 553 SVSSSFISSSESKSP---TYSSSLPLVTISATTSQETASSL-PPATTKTSQTTILVT- 607
QY 583 YALSNCTFYRSTDDGGVTFQPVAAAGLPSGAVGMVHAPV--KEGDLWLAASS----- 633
Db 608 -----VTSCESHVCTESISPAIVSTATV-----TVSGVTTEYTTWCPITSTETTKQT 654
QY 634 -GLYHSTNGSSWSAITGYSSAVNNGVFGKSAPGSSYPAVE--VVGITGGVTAIRSDDCG 690
Db 655 KGTTEQTETTTKQTTVTVTISSESDVCSTAS-----PAIVSTSTATINGVTTEY----- 704
```


QY 691 TTWVLIN 697
| | | | |
Db 705 TTWCPIS 711

RESULT 9

US-08-971-692-15

; Sequence 15, Application US/08971692
; Patent No. 6114147
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immobilized proteins with specific binding
; TITLE OF INVENTION: capacities and their use in processes and products.
; NUMBER OF SEQUENCES: 40
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,692
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-971-692-15

Query Match 3.5%; Score 143; DB 3; Length 894;
Best Local Similarity 20.5%; Pred. No. 0.023;
Matches 161; Conservative 94; Mismatches 308; Indels 224; Gaps 39;

QY 7 TWSNVAIGGGFVDGIVFNEGAPCILVVRTDIGMYRWDAANGRWIPLLDWVG----- 59
| | | | | : : : : : | | | | | : : : : :
Db 53 TYSNAAVWAGYA-----SKTKLGVGGQDHSIDYN-----IPCVSSTGTFPCQE 99
| | | | | : : : : : | | | | | : : : : :
QY 60 --WNNWNGYGVSTADPINTKV--WAA--VGMYTNSWD----- 93
| | | | | : : : : : | | | | | : : : : :
Db 100 DSYGNWCKMGAC-----NSQGIAYWSTDLFGFYTPNTVLEMTGYELPQGTGYTFK 155
| | | | | : : : : : | | | | | : : : : :
QY 94 --PNDGAILRSSDQATW-----QITPLPFKLGNNMGRGMRGLAVDPNNDNII 141
| | | | | : : : : : | | | | | : : : : :
Db 156 FATVDDSAII--SVGGATAFNCCAQQOQPPITSTNFTIDGKPMGG-----SLPPNIEGV 208
| | | | | : : : : : | | | | | : : : : :
QY 142 YFGA---PSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWAFDK 197
| | | | | : : : : : | | | | | : : : : :
Db 209 YWAGYYPW-KVYSNVSWGILPISVTLPD-GTIVSD-----DEGYVY-SFD- 255
| | | | | : : : : : | | | | | : : : : :
QY 198 SSSSLGQASKTIFGVADPNPNVFWSRDGGATWQVPGAPTGFIPHKGVDPVNHVLYIA 257
| | | | | : : : : : | | | | | : : : : :
Db 256 --DDLQSQNST--VDPDSN-----YAVSTTTTTEPWTGTSTSTEMTV 297
| | | | | : : : : : | | | | | : : : : :
QY 258 TSNTGGPYD-----GSGDWKVFSTGTWTRISPVPTDNDYFGYSLGLTID-- 306
| | | | | : : : : : | | | | | : : : : :
Db 298 TGTNGVPTDETIVIRPTSEGLI---STTEPWTGTFTSTSTEVTT-ITGTNGQPTDET 353
| | | | | : : : : : | | | | | : : : : :
QY 307 ----ROHPNTIMVATQISWMPDTIIFRSTDGATWTRIDWTSYPNRSLRYVL----- 355
| | | | | : : : : : | | | | | : : : : :
Db 354 VIVIRPTSEGLISTTTEPWTGFTSTSE---MTVTGTNGQPTDETIVIRPTSEG 409
| | | | | : : : : : | | | | | : : : : :
QY 356 --DISAEPWL-TRGVQPNPVPVSKLGWDEMAIDPFNSDRMLYGTGATLYATNDTKW 412
| | | | | : : : : : | | | | | : : : : :
Db 410 LVTTTTEPWTGTF-----TSTSTENSTVGTNGLPDTEV-- 444
| | | | | : : : : : | | | | | : : : : :
QY 413 DSGQIHIAPVWKLEETAVNDLISPSPA-----PLISALGDLGFFHADYAV 462
| | | | | : : : : : | | | | | : : : : :
Db 445 -----IVVKTPPTTALSSSLSSSSGQITSTTSRPIITPFYPSNG-TSVISSV 493
| | | | | : : : : : | | | | | : : : : :

QY 463 PSTFTSPVETTGTSDYAEINPSIIVRAGSFDPSOPNDRHVAFTSDGKNWFQSGEPG 522
| | | | | : : : : : | | | | | : : : : :
Db 494 ISSSVTSSLTSPFVSSSVISSSTTTSTSTIFSESSKSSVPTSSSTSGSSE-SETSSAG 552
| | | | | : : : : : | | | | | : : : : :
QY 523 GVTGTGTAASADGSRFVWAPGDPGVVYAVGFGNSWAAASQGVPAQAQIRSDRVNPKTF 582
| | | | | : : : : : | | | | | : : : : :
Db 553 SVSSSFISSESSKSP---TYSSSLPLVTSATTSQETASSL-PPATTTTSEQTTLVT- 607
| | | | | : : : : : | | | | | : : : : :
QY 583 YALSNCTFYRSTGDTGTFQPVAAAGLPSSGAVGVNFHAVPG--KEGDLWLAASS----- 633
| | | | | : : : : : | | | | | : : : : :
Db 608 -----VTSCESHVCTESISPAIVSTATV-----TVSGVTTEYTTWCPISTTETTKQT 654
| | | | | : : : : : | | | | | : : : : :
QY 634 -GLYHSTNGSSSAITGVSSAVNVGFKSAPGSSYPAVF--VVGTTGGVTGAYRSDDCG 690
| | | | | : : : : : | | | | | : : : : :
Db 655 KGTEQTETTKOTTVTVTISSCESDVCSTAS---PAIVSTSTATINGVTTEY----- 704
| | | | | : : : : : | | | | | : : : : :
QY 691 TTWVLIN 697
| | | | |
Db 705 TTWCPIS 711
| | | | |

RESULT 10

US-09-296-284-25

; Sequence 25, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-25

Query Match 3.5%; Score 141; DB 4; Length 720;

Best Local Similarity 20.4%; Pred. No. 0.024;
Matches 145; Conservative 73; Mismatches 235; Indels 258; Gaps 43;

QY 26 EGAP---GILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVV--SIA----ADP 75
| | | | | : : : : : | | | | | : : : : :
Db 58 EGTPLIVDGVMYATTNWSKMKALDAATGKLL-----WSYDPKVPGNIDRGCCDT 107
| | | | | : : : : : | | | | | : : : : :
QY 76 INTNKVWAAVCMYTNWDPNNDGAI--LRSSDQATWQITPLPKLGNMGRGMRGLAV 133
| | | | | : : : : : | | | | | : : : : :
Db 108 VNRGAAYWNGKVYFGTGT---DGRLLIALDAKTGLVNSVYTVP-----KEAOLGHQRSYTV 159
| | | | | : : : : : | | | | | : : : : :
QY 134 DPNNNDNILEYFAGPS--GKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVW 192
| | | | | : : : : : | | | | | : : : : :
Db 160 D-----GAPRIAKGVIIIGNGAEGF-ARCF-----VTAYDAETGKMDW 197
| | | | | : : : : : | | | | | : : : : :
QY 193 VAF---DKSSSLGQASKTIFGVADPNPNVFWSRDGGATWQVPGAPTGFIPHKGVDP 249
| | | | | : : : : : | | | | | : : : : :
Db 198 RFTVPNPNKPDGAAASDDVLMKAYPT---WKG--GGAWKQQGGG--GTVWDSLIIYDP 249
| | | | | : : : : : | | | | | : : : : :
QY 250 VNHVLIATNTGPGYDGGSDGVWKFSTVSGTWTTRISPVSTDTANDYFGYSLGLTIDRQH 309
| | | | | : : : : : | | | | | : : : : :
Db 250 VTDLVLVGVGN-GSPWN-----YKFR-----SEKGNLFLGSIVAIN--- 286
| | | | | : : : : : | | | | | : : : : :
QY 310 PNTIMVATQISWMPDTIIFRSTDGATW-----TRIDWTSYV-----NRSURYV 354
| | | | | : : : : : | | | | | : : : : :
Db 287 -----PDT-----GKYVWHFQETPMDQWDITSVQOIMALDPVNGEMRHV 326
| | | | | : : : : : | | | | | : : : : :
QY 355 LDISAEPWLTFGVQPNPVPVSKPLGWNDEMAIDPFNSDRMLYGTGATLYATNDLTKWDS 414
| | | | | : : : : : | | | | | : : : : :

Db 327 L-----VHAPKNGFF---YIIDA-KTGKFIISKPKVT-----YENWAN 359
QY 415 G-----GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDGLGFTHADVTAVPSTIFTS 469
Db 360 GLDPVTRGRPNP-----DALWTLNGRPWYIPGDLGCHNFANAYSPQ---TK 405
QY 470 PVFTTGTSDYAEALNPSIIVRAGSFDSPSQPNDRHVAFTDGGKNWFGQSEPG--GVTTG 527
Db 406 LVYIPAQQVFPV-----YDPKGGFKAH-----HDSWNLGLDMKMKIGLLDD 446
QY 528 GTVAASADGSRFV-----WAPGDPGPVYVAVGFGNSWAASQGVAPANAQIRSDRVN 578
Db 447 NDPQHKADKAQFLKDLKGWIVAMPD--QKQAAFTVDHKGPM-----486
QY 579 PKTFYALSNGTFFRSTDDGVTTFQPVAAAGLPSGAGVGMFHAVPGKEG-DLW-LAASGLY 636
Db 487 -----NGLL-ATAGGVLFQGLANG-----EFHAYDATTKDKLFTTFAQSAIL 528
QY 637 -----HSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVVGTIGGVTG 682
Db 529 APPVTYTANGKQ-----YVAVEGW-----GGIYP--FFLGGVARTSG 564
RESULT 11
US-09-296-284-4
; Sequence 4, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-4

Query Match 3.5%; Score 141; DB 4; Length 754;
Best Local Similarity 20.4%; Pred. No. 0.026;
Matches 145; Conservative 73; Mismatches 235; Indels 258; Gaps 43;

QY 26 EGAP-----GILYVRTDIGMYRMDAANGRIPLLDWYGNNGYGVW--SIA-----ADP 75
Db 92 EGTPLIIVDGVYATTNWSKMKALDAATGKLL-----WSYDPKVPGNADRGCCDT 141
QY 76 INTNKVAAVGMTNSWDNDGAI--LRSSDQAGTWOITPLPFKLGNNMFGRMGERLAV 133
Db 142 VNREGAATWNGKVFEGT---DGRLLALDAKTGKLVMSVTVTP-----KEAQLGHQRSYTV 193
QY 134 DPNNNDILYFGAPS-GKGLWRSTDGATWSQMTNEPDVGTYIANPTDTTCGYQSDIOGVVW 192
Db 194 D-----GAPRIAKGVIIINGGAIEFG-ARGF-----VTAIDAEKMDW 231
QY 193 VAF---DKSSSLGQASAKTIFGVADPNPNPVFWSRDGATWQAVPGAPTGFIPHKGVFDP 249
Db 232 REFTVPNPNKPDGAASDDVLMSKAYPT---WGK--GGAWKQGGG--GTWVDSLLIYDP 293
QY 250 VNHVLIATNTGPGDVGSSGDVWKSFSVTGWTTRISPVSTDTANDYFGYSGLTIDRQH 309
Db 284 VTDLVVLGVGN-GSPWN-----YKFR-----SEKGNMFLGSIIVAIN---320
QY 310 PNTIMVATQISWMPDRIIFRSTDDGATW-----TRIDWMTSYN-----NRSIRVY 354
Db 321 -----PDT-----GKYVWHFQETPMQWDITTSYQQIIMALDMPVNGEMRHV 360

QY 355 LDISAEPWLTFTGVQPNPPVPSPKLGMWDEAMAIIDPNFSDRMILYGTGATLYATNDLTKWDS 414
Db 361 L-----VHAPKNGFF---YIIDA-KTGKFIISKPKVT-----YENWAN 393
QY 415 G-----GQIHIAPMVKGLEETAVNDLISPPSGAPLISALGDGLGFTHADVTAVPSTIFTS 469
Db 394 GLDPVTRGRPNP-----DALWTLNGRPWYIPGDLGCHNFANAYSPQ---TK 439
QY 470 PVFTTGTSDYAEALNPSIIVRAGSFDSPSQPNDRHVAFTDGGKNWFGQSEPG--GVTTG 527
Db 440 LVYIPAQQVFPV-----YDPKGGFKAH-----HDSWNLGLDMKMKIGLLDD 480
QY 528 GTVAASADGSRFV-----WAPGDPGPVYVAVGFGNSWAASQGVAPANAQIRSDRVN 578
Db 481 NDPQHKADKAQFLKDLKGWIVAMPD--QKQAAFTVDHKGPM-----520
QY 579 PKTFYALSNGTFFRSTDDGVTTFQPVAAAGLPSGAGVGMFHAVPGKEG-DLW-LAASGLY 636
Db 521 -----NGLL-ATAGGVLFQGLANG-----EFHAYDATTKDKLFTTFAQSAIL 562
QY 637 -----HSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVVGTIGGVTG 682
Db 563 APPVTYTANGKQ-----YVAVEGW-----GGIYP--FFLGGVARTSG 598
RESULT 12
US-09-352-159-25
; Sequence 25, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 1000
; TYPE: PRT
; ORGANISM: Unknown
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-25

Query Match 3.5%; Score 140.5; DB 4; Length 1000;
Best Local Similarity 20.9%; Pred. No. 0.041;
Matches 159; Conservative 69; Mismatches 224; Indels 309; Gaps 41;

QY 150 GLWRSTDGATWSQMTNEPDVGTYIANPTDTTCGYQSDIOGVVWVAFDKSSSLGQASAKTI 209
Db 16 GLSASLASGAPT VKI---DAGMVVGTITVPGTTATVSEFLGVFP-----AASPTR 63
QY 210 FVGVADPNPNPVFWSRDGATWQAVPGAPTGFIPHKGVFDPVNHVLIATNTGPGYDGS 269
Db 64 F---APTRFVPWSTPLQATAYG-PACPOF-----NYPEELREITWAMFNTPPPAGES 114
QY 270 GDWKFVSFTSGWTTRISPVSTDTANDYFGYSGLTIDRQHPTIMVATQISWMPDIIIFR 329
Db 115 EDCLNLNIY-----VPGTENTN-----KAVWV-----W-----IY- 139
QY 330 STDGGAT---WT---RIWDWTSYPNR-----SLRYVLDISAEPWLTFTGVQPNPPVSPKL 378
Db 140 ---GGALEYGWSFHLVDGASFAANDVIATVINTYNTNI-----LGFPAPAPLPITQRLN 191

RESULT 15

```

RESULT 15
US-09-352-159-29
; Sequence 29, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gulliam, Joyce R.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyl Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 1205
; TYPE: PRT
; ORGANISM: Unknown
US-09-352-159-29

```

Query Match 3.4%; Score 135.5; DB 4; Length 1205;
Best Local Similarity 21.3%; Pred. No. 0.12;
Matches 176; Conservative 71; Mismatches 230; Indels 349; Gaps 49;

Qy	94	PNDGAILRSSD-----QGATWITPLPFLKGLGNMPGRGMBERLAWDPNDNDILYFGAPS	147
Db	187	PQIDKYLKSSKYIAWPLQG--WQAT---FG-GGDHP-----PKSDLV-----PR	224
Qy	148	GKGLWRST---DSGATWSQMTNEPDGYTIANPTDTGTYSQSDLOGVVWVAFDKSSSLQG	204
Db	225	GSPEFAPTVKIDAGV-----VGTITTVF-GTTATVSEFLGVPPFA-----	263
Qy	205	ASKTIFGVGADPNPNFWSRDGATWQAVPGAPTGTIPHKVGFDPVNHVLYIATSNITGGP	264
Db	264	ASPTRF---APPTRPVWSTPLQATAYG-PACPQQF-----NYPEELREITWAWENTPPP	314
Qy	265	YDSSSGDVWKFSTVGTWTRISVPSTSDTANDYFGYGLGTLDRQHPNTIMVATQISWMPD	324
Db	315	SAGESEDECLNLT-----VPCGTENTN-----KAVWV-----W----	342

Query Match 3.48: score 136: DB 2: Length 2628:

every match 3.4%; Score 150, SS 2, length 2026,
Best Local Similarity 20.1%; Pred. No. 0.33;
Matches 178; Conservative 68; Mismatches 265; Indels 374; Gaps 52;

Qy	69	VSTAADPIWTKV----	WAAVQM--YTNSWD-----	PNDGAILURSD-----	104
Dq	70	:::	:::: :::: ::::	:::	
Db	917	VNYTVDFOVFNPNQNITGS	SAVGKQVTLKWDA	PNGTNPNNPNCPTTTLSSEFENGIPAS	976
Qy	105	-----OGATWOITPLFKLGGNMPGRGMERLAV-	-GGT---SFGHNSAICASSASYINFEPPONPNYLVT	--DNNENILY- -	142
Db	977	WKTIADGDGNWMTTPPPP	-----GGT---	SFGHNSAICASSASYINFEPPONPNYLVT	1030
Qy	143	-FGAPSGKGL--W-----	RSTDGSAQTOMTFPD-----	-VGTVIANP--	177
Db	1031	ELSLPNGGTLTFMVCQAODANYASEHYIVAYSTGDNDSANEALLLEVLTAKTVVTAPEA	:-: : : :-: : : :-: : : :-: : : :-: : :		1090
Qy	178	----"-TDTTYQSDDIQ--GVNVAFD-----	KSSSLCOAQSKTIFVG	212	
Db	1091	IRCTRQGTWYQKTQLGPAGTYIAFRHFRCGDTDFEWINLDVEIKANGKRADFET-FES	1149		
Qy	213	VADPNPNFWIS---RDG-CATWQAVPGAPTGFPHKG-----	VFDPVNVL- 254		
Db	1150	STGEPAEAETHIDADCQGHLCSGOLGWLTAGHTNVVASFSWGMAINFNLIS	1209		
Qy	255	-----YIATSNTGGPYDGSSGDVKFVSJTGTRISPVPSTDANDY-----	F 298		
Db	1210	KDVTGATKVKYVAVNGFPDGHYAMISKGTNAGDET---	VVFETPNINKGGARF	1265	
Qy	299	GYSGLIIDRQHPTNWATOISWDPDTIIFRSTDCGATWTRIWDWTSYPNRSRYLV--D	356		
Db	1266	GLS--TEADGAKPQSV-----W----	IERTVDLP-A-GTKYVAFRHYNCSDLNIYLLDD	1311	
Qy	357	ISAEPWLTFGVOPNPVPSPKLGWDEAMA IDPFNSDRMLXGTGATLYATNDLNKNWDSGG	416		
Db	1312	IQ-----FTMGGSTP-----	TDYTYVYRGDKRK-----	1335	
Qy	417	QIHIAPMWKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSPTIFTSPVTTGT	476		


```

QY 325 TIIFRSTDCGAT---WT---RIWDWTSYPNR-----SLRYVLDISAEPLTFGVQVQPPV 373
Db 343 --IY-----GGALEYGWNSEHLYDGASFAANQDVIAVTINRTNI-----LGFPAAFPOLPI 391
QY 374 PSPKLGWMDAMAID-----PENSOR 394
Db 392 TORNLGFLDQRFALDWQORNIAAFGDPKRVTFGQSAGRSVDVLLTSMPHNPPFRAAI 451
QY 395 MLYGTGATLYATNDLTK-WDSGGQIHIAPMVKGLEETAVNDLIS---PPSGAPLISALGD 450
Db 452 MESGVANYNFPKGDLEPWN-----TVQALNCTTSIDILSCMRVRDLATIMNTIEQ 503
QY 451 LG-GFTHA--DVTAVPSTFTSPVFTGTSVDYAEIENPSTIIVRAGSFDPSOPNDRHVA 507
Db 504 LGLGFEYTLDNVTAV---YRSE--TARTGDIARV-PVLV-----537
QY 508 STDGKNWFQSGPEPGVTGTTVAASADGSRFVWAPGD-----PGQPVVYAYVFGN 558
Db 538 -----GTVA--NDGLLFLVGENDTQAYLEEAIPNOPDLYQTLGA 575
QY 559 SWAASQGV--PAN-----AQIRSDRVNPKTFYALSNGTF-----590
Db 576 YPIGSPGIGSPQDQIAAIEFEVRFQCPISAIVAODSRNRGIPSWRY-YYNATFENLELFP 634
QY 591 ---YRSTDGCVTF--QPVAAG-----LPSSGAVGV 615
Db 635 SEVYHSEVGMVEGTYTPVASATALEAQTSKYMOGANAFAKNPMNGPGWKQVNPVAAALG- 693
QY 616 MFHAVPGKE-----GDLWLAASSGLYHSTNGSSWSAITGVSSA 654
Db 694 -----SPGKAIQVDVSPATIDQRCALYTRYVTELTGTPRTFGGGGGGGGSKDNVADV 749
QY 655 VNVGFGKS-----APGSSYPAVFVVTIGGVTGAYRS-----DDCGTTWVLND 698
Db 750 VVVGAGLSLETARKVOAAGSLCVLEAMDRVGGKTLVQSGPGRTTINDLGAW--IND 807
QY 699 DQ-----HOYGNMGQAITGDHANLRVYIGTNGRGIVYGD 733
Db 808 SNQSEVSRUFERFHELEGEL-ORTTGN--SIHQADGCTTTTA-PYGD 849

```

Search completed: July 2, 2002, 09:12:54
Job time: 62 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:13:57 ; Search time 54.67 Seconds
(without alignments)
1300.642 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YICTNGRCIVYDGGAPSG 740
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97013	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical prote
6	202.5	5.0	2174	2 E95965	hypothetical glyci
7	189	4.7	3972	2 S75251	hypothetical prote
8	186.5	4.6	4199	2 S78412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RFX famil
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 AB3528	extracellular seri
17	166.5	4.1	902	2 H87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1884	2 S10789	amylase A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD0835	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

30	157	3.9	809	2 A55547	quinate-shikimate
31	156.5	3.9	2204	2 A70524	probable PPE prote
32	156	3.9	3157	2 B70969	probable PPE prote
33	155	3.8	4180	2 G83559	hypothetical prote
34	154.5	3.8	1032	2 T34433	hypothetical prote
35	154.5	3.8	3570	2 T45025	mucin MUC5B, trach
36	153	3.8	1055	2 A87364	OmpA-related prote
37	133	3.8	1821	2 AG2335	hypothetical prote
38	153	3.8	13288	2 T03099	mucin, submaxillar
39	152.5	3.8	3016	2 S77300	hypothetical prote
40	152	3.8	348	2 T35248	probable oxidoredu
41	152	3.8	1049	2 T42045	beta transducin-li
42	152	3.8	1196	2 A29130	beta-amylase (EC 3
43	152	3.8	3472	2 T31308	hypothetical prote
44	151.5	3.8	820	2 B72575	hypothetical prote
45	151.5	3.8	1145	2 B75625	hypothetical prote

ALIGNMENTS

RESULT 1

T35237

probable secreted cellulase - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C;Accession: T35237

R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z21572

A;Accession: T35237

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-890 <SEE>

A;Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GNO00070; SCOEDB:SC5C7.30C

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC5C7.30C

Query Match 58.7%; Score 2371; DB 2; Length 890;

Best Local Similarity 57.7%; Pred. No. 1.3e-131;

Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY 6 YTSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWITPLLDWVGNNWGY 65

DB 41 YTKNARIDGGFVPGIVFNRTEKDLAYARTDIGGAYRWQESHTWPTLLDHVGWDDWGH 100

QY 66 NGVYSTAADPINTNKYWAAGMYTNSWDNDGATLRSSDOGATWQITPLPKLGGNNMGR 125

DB 101 TGVALASDAVDPRVAAVGTNTNDMDPTNGAVLRSDRGASWEKADLPKLGNNMGR 160

QY 126 GMGERLAVDPNNNIIYFGAPSGKGLWRSTDGATNSQMTNFPDVGTYIANPTDTTCYQS 185

DB 161 GMGERLAVDPHNDVLYLGLAPSGHGLWRSTDAGVTNSEVTAFFPNPNYADPNDSYAS 220

QY 186 DIQGVVWVAEDKSS-SSLGASKTIFVGADPNPNVFWSRDGGATWQAVPGAPGFTPHK 244

DB 221 DNQGITWVTDESTGGAGTATRLVYGADKENAVYRSTAGATWERLAGQPTGYLAHK 280

QY 245 GVDPDNNHLYIATSTGGPYDSSGDVWKFVSTGTTWTRISPVSPSTDANDFGYSGLT 304

DB 281 GVLDANGYLYLAYSPTGGPYDGGKGLRYATATGTWTDISPAEAADT---YFGSLT 337

QY 305 IDROHNTINVAQTQISWNPDTIIFRSTDGATWTRWDWTSYPNRSRLRYVLDISAEPLWT 364

DB 338 VDRQRGCTVMATAYSSWWPDTQIFRSTDGATWQASWYSYSDRENRYTMDVSSSWLT 397

QY 365 FGQVQPNPPVSPKLGWDEAMADPPNSDRMLYGTGATLYATNDLTKW-DSGGQIHIAPM 423

DB 398 WGANPAPPEQTPKLGWTEALEIDPFDSDRMYGTGATVYGTENLTWNDDDEGGTFAVEPM 457

QY 424 VKGLEETAVNDLISPPSGAPLISALDGLGFTHADVTAVFSTIFTSPVFTTGSVDYAE 483

Qy	362	WLTFG - YOPNPPVPSPKLGWMD	EAMADP	FNDSRMLYGTGAT	YATYNDLTKWDSGG	OIH	420																																																	
Db	389	WLDWGTGVT	PPDP	PLVFLGWMKMG	DLEDP	FNDSRMEFYGTGAT	LYGTDDLTNWDKGNVDI	448																																																
Qy	421	APVYKGL	EEETAVNDL	ISPP	SCAPLISAL	GDGLGFT	HADVTAVPSTFT	SPVFTTGT	SVDY	480																																														
Db	449	SVKANG	TEEC	AVNDVV	YPTT	GAQLLS	AVGDCGFY	DDIT	KVP	KMMTTFNF	SATTSIDY	508																																												
Qy	481	AELNPS	IT	IVRAGS	PDPS	QPNDR	HVAF	STD	CGKNWFO - GSEP	GGVT	TGGT	VVAASADGSRF	539																																											
Db	509	AESVP	NFVVR	GVN	VDTS	SKNQ	ODKDCG	ISY	DGKNWFS	AGS	NI	SGVYKAGT	VAAGADAKTI	568																																										
Qy	540	VWAF	GD	PGQ	PVY	YAV	FGNS	WAA	SQGV	PANAQ	IRSD	RNP	KPTFYA	LSNGT	FYRST	DGGVY	599																																							
Db	569	WVSP - EEC	ANAA	ST	DNG	NKWT	PC	SLP	Q	QAC	VRSD	RNP	NPKFY	GF	LNG	RY	IS	DAGAT	627																																					
Qy	600	F - QP	V	AGL	PSS	G	AVG	YMF	H	V	PCKE	GD	LWLA - ASS	GLY	H	ST	NGG	SS	W	SAIT	GVSS	AVNV	657																																	
Db	628	FTQ	SSQ	T	GLP	T	KGK - G	I - F	K	T	V	I	G	H	E	G	D	I	W	I	A	G	K	D	L	W	H	S	T	D	G	A	T	F	K	V	S	G	V	D	A	S	T	V	685											
Qy	658	GFGK	S	A	P	S	S	S	S	P	A	V	F	V	V	Y	T	G	G	V	T	G	A	V	R	S	D	D	C	G	T	T	W	L	I	N	D	D	O	H	Q	Y	G	N	M	G	O	A	T	T	G	H	A	N	L	717

```

Qy      718  RRVYIGTNGRGIVYGDIGGA 737
      ||:|||||
Db      746  GRVFGTNGRGIVYGDIDGS 765

RESULT      3
T00349
Avicelase III - Aspergillus aculeatus
C/Species: Aspergillus aculeatus
C/Date: 01-Feb-1999 #sequence: T00349
C/Accession: T00349
R/Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
submitted to the EMBL Data Library, June 1998
A/Description: Avicelase III from Aspergillus aculeatus.
A/Reference number: Z14141
A/Accession: T00349
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-856 <ARA>

```

C;Gene: avIII
C;Superfamily: fungal cellulose-binding domain homology
F:823-854/Domain: fungal cellulose-binding domain homology <FCR>

Query Match	41.6%;	Score 1680;	DB 2;	Length 856;
Best Local Similarity	46.2%;	Pred. No. 4.8e-91;		
Matches 346;	Conservative 113;	Mismatches 258;	Indels 32;	Gaps
Qy	1	ATTQPTYSNVAI--GGGGFVDGIVFNEGAPGILYVRTDIGGMYRWDAAANGRWIPLDLWDVG	50	
Db	21	AASQAYTKNVVTGGGGFTPGIVFNPSAKVAYARTDIGGAYRLN--SDDTWIPLMDWVG	79	
Qy	60	---WNNNGVNGVSIADPINTNKKVAAQVMYNSMDPNDGAILRSSDQCATWQITPLPF	116	
Db	80	NDTWHDW---GIDALATDPVTDTRVYVAVGMYTNEMDPNVGSIILRSTDQDGTWTETKLPF	136	
Qy	117	KLGNMPPGRGMRGLAVDPNNDNILLFPGASGKGLRSTDSGATWQSMTNFPDVGTYIAN	176	
Db	137	KVGGMPPGRGMRGLAVDPNKNLSILFEGARSGHGLMKSTDYGATSNVTSFTWTGYFQD	196	
Qy	177	PTDFTGYQSDIQGVVVAFFKSSSSILGOASKTTFVGVADPNPNVFNRSRGGATWQAVPGA	236	
Db	197	SSST--YTSDPVGIATWTFDSTSGSSGSATPRIFGVADAGACKSVFKSEDGATWAVSGE	254	
Qy	237	PT-GFIPHKGVDPVNNHVLVIATSNTPGGPYDGGSSGDVWKFESVTSQWTRISPVSPSTDAN	295	

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75251
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S75251
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3972 <KAN>
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAA17165.1; PID:g165224
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr1028

Query Match 4.7%; Score 189; DB 2: Length 3972;
Best Local Similarity 19.3%; Pred. No. 0.013;
Matches 166; Conservative 93; Mismatches 276; Indels 326; Gaps 42;

QY 13 IGGGGFVGVFNEGAPGILYVRTDIGMYRMDAANG-----RWIPLLDW- 57
Db 2458 VNGDGFADVISGGSPAGGVLIIFGNTKDLL--DAALGTDLIIISVENAQVKEFVALGDFD 2515
QY 58 -----VGNWNGYNG---VVSTAADPINTNKVVAAGMTNSWD-- 93
Db 2516 GDGLADFGVDDGNGFLLVGLSGPQLSLVLDSTLPNLNFAQWGVDFNGYDFF 2575
QY 94 ---PNDGAILRSDDCATQWITPLPKLGNMP-----GRGMCERLAVDPN-- 136
Db 2576 VLOGPNSTIAYGNANGTLTDSPLTF--GNPFPLSSFTGIDLNGNKEIVAGQPNLN 2633
QY 137 -----NDNIIYFGAPSGKGLWRST-----DSG-ATWSQMTNFPDVGTYIANP 177
Db 2634 PVPNIGFGGGLGYFTYEAGNAVLPQTPNPNASVTEASGLSWGQI-SPPNQYAAQAVP 2692
QY 178 TDRT--GY-----QSDIQGVVW-----VAFDKSS----- 200
Db 2693 SFATLGLWLYQAFYGINERISTKDSYIYIORSRDSGSWENLTQVPLDSNGTPIDLKNLP 2752
QY 201 -SLGQASKTIFVGVADPNPNPVFWSRDG-----GATWQAVP-----GAPT--GFTPHK 244
Db 2753 PSITAYNGTLYLFTADNGQV--WVAEGVNTNANSGLILNAVPIQNASNGPTLVAFNDEL 2811
QY 245 GVF-----DPNVHLYIATSWTGPYDGSQGVKFSVTSCTWTRISPVSTDRANDYFGY 300
Db 2812 YVFFVKDASNDILYSSSSNPG-----SSSG--WDGTSTLTFTSDVNOATN-----FPL 2858
QY 301 SGLTIDRQHPNTIMWATQISWMPDTIIFRSTDGATWTRI-----WD----- 342
Db 2859 SATVVPGLDGTILAVA-----FRNNSPATWVGLNLSDDVTNWOGSBELTQVDA 2907
QY 343 -----WTSYPNRSRLRYL-----DISAEPWLTFGVQPNPVPSPKL 378
Db 2908 NSQVSLTVVDGTYLYLFTSTSEASVATSTDLGNWGDITLIPW-----DDGNL 2956
QY 379 GWME-----AMADPFNSDRMLYGTGATLYATNDLTKWDSGGQTH----- 419
Db 2957 GGVAIFLNFQSFILSLNQSNNESLLFAFSNSLFEPNQASRW--GEQVRDIFDNGDGIAD 3014
QY 420 ---IAPVKGLETAVNDLISPPSGAPLISALGDL-----GGFTHADVAVPSTIFTSPVF 472
Db 3015 LAVIAPGYRNLLQFPILDY-----PAINNLGVFTIYGEESGISVNDPDPVLAAPDL 3067
QY 473 TTGTSVDYAEINSLIVRAGSDP--SSQP-----NDRHVAFTDGGKNW-- 515
Db 3068 PQETIFELLEITPTGDVNGDGFDDLLISAPLTPVIAQGFDPVNGDGVSVWVFGGTHWG 3127
QY 516 -FOGSEP-----GGVTG-----GTVAAASADG--SRF-VW 541
Db 3128 EYTANSPFGLGNLANNQTNNSQNFYGFVTTLGPRSQAGISISGGADVNGDGFSDFALG 3187

QY 542 APGD-----PGQPVYAVGFGNSWAASQG---VPANAQTRSDR 576
Db 3188 APGNFNLNLYVLFSGSFTNOVNLGTIGDDVMLGSPTEIFVAGQGDQIYTTNGGVDIVY 3247
QY 577 VNPKTFYALSNGTFFYRSTGSG 597
Db 3248 AGPNDFFVVTDTNFRRLDGG 3268

RESULT 8
S76412
hypothetical protein.slr0408 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76412
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76412
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4199 <KAN>
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAA18541.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr-0408

Query Match 4.6%; Score 186.5; DB 2: Length 4199;
Best Local Similarity 19.2%; Pred. No. 0.019;
Matches 198; Conservative 124; Mismatches 319; Indels 389; Gaps 54;

QY 9 SNVAIGGGGVGDIV-----FNEGAPGILYVRTDIGMYRMDAANGRWIPLL 55
Db 2261 ADVASDNGFVIDGNLIGNPPTTETTSOYIDTTPAILINGSNLYLAYKFGGNNQ---IY 2317
QY 56 DWYGVNN--MGYNGVWSTAADPINTNKV--AAGVMYNSW-----DPNDGA-ILRSSDQGA 107
Db 2318 FTVSTNNGOSWSEVOL---PQSAQTIFPPAIAFFNNVLYAVDGNNGNLNITSDQGG 2374
QY 108 TWQITPLPKLGGNMPG-----RG-MGERLAVDPNNDNILEGAPSGKGLRSTDSGA 159
Db 2375 TWN---APLALGTSSTPPPLFVYQGLTLLFAANNSTSTVQFYLNSSNEWIYANEIGS 2431
QY 160 TWSQMT-----NFPDVGTYIANPTDTTGYQSDIQGVWVAF--- 195
Db 2432 NQTAISATVLDGTLILYVYKGGTRNTPTLDYITSTN-----ADLSANDWSSPIPG 2486
QY 196 --DKSSSSLGQASKTIFVGVADPNPNPV-FWSRDGGATWQAVPGAPTGF---PHKGVFDP 249
Db 2487 VSSQGGPSLTNDGTNLVLSYLDSSNQLNFVSSGNGINWSS-POVITNINISQSPAIATF-- 2543
QY 250 VNHVLIATSNCTGPGDGGSDGVKFSVTSCTWTRISVPST-----DTAN 295
Db 2544 ANNELYLS-----YPGQGS-QELNVTs-----FPLPFTGSILNGSLVRLGVDNG 2589
QY 296 DYFG--YSGLTIDRQHPNTIM-----VATQISWMPDTIIFRSTDGATWTRIDWTSY 346
Db 2590 DGFADVFSGGT---NAGALIFGNSTKDLTTTASGSEDLVI-----SV 2628
QY 347 PNRSLRYVL---DISAEPWLTFGV-----QPNPVPVSPKL 378
Db 2629 PNATLRDVTSGVDFNGDGIKDLGLDNGNFYVVLGNTSLGLDKLTLSITSSSPVIVNOV 2688
QY 379 GWDEAMADIPFNSDR---MLYGT-----GATLYATNDLTKWD----- 413
Db 2689 GGVTKSAIGDYGNGDYDDVLLMGDNGCNQVAGNSGTGLNSFTNIDYPTETQTTATGVDLN 2748
QY 414 -----SGGQIHAPVMVKGLEETAVNDLISPPSGAPLISALGDLGGFTHAD 458

Db 2749 SDGIPEIAIGDERKIAQIOISTSGSFLPTTSSVINTLAAANOLENIGDFNGCIAD 2808
QY 459 VTAVPSFTIFS-----PVFTTGTSDVYAEIN-----PSLIIVRAGSFDPS 497
Db 2809 LAVLASWYAAAICEPNLNLPSRPGNOGGVFIFYGNSNGLSNTAQPDVILAAPTNPS 2868
QY 498 SQ-----PNDHRHAFSTDGKNW----- 515
Db 2869 GQISTYOLSRIAQAGDVNGDFDILLISSPYTDAENNOQGVFVFGDDRNQPFDLGQ 2928
QY 516 -----FGSEPPGGVTGTVAASA-----DGSRF-----VWAPG----- 544
Db 2929 LRANQSGSNPRFAIDGSPNSOAGIALNGGDIINGDFADFIIGAPGENNLQYNQOIVF 2988
QY 545 -----DPGQPVYAVG-----FGNSWAASQGVPAQAQIRSDRYN-----PKTFYAL 595
Db 2989 IENGELSDDDKYSYILYLDGNOTOMGGGDWQANO-VWNT-QVATWNNSRPRPEAVIGQ 3046
QY 586 SNGTFYRSTDGGVTFQ-----PVAAGLPSSGAVGVMFHFAVPGK 623
Db 3047 SNGDIWYPCGNQWQSWGKLPAEINELAVNNTSGNPQIIAGLGKG--GIEYY----- 3099
QY 624 EGDWLAAASGLYHSTNGSSW-SAITGVSSAVNVGFGKSAPGSSYPVAVFVGTIGVGTG 682
Db 3100 NGSTWV--NNGPYQ---GDCWRSAITQMA---VOMGED--GS--PSQIVVGLADCAVI 3145
QY 683 AYRSDDCGTTWVLNDQHOYGNWGOAITG-----DHANLRVYIGTN 725
Db 3146 YYNTQ---SCWRTIN-----NFGKSVTOLSVMQEAASNPVIVGLDNSEV-QYQGSN 3194
QY 726 GRGIVYGDIG 735
Db 3195 GVWTFPHDDG 3204

RESULT 9
T13256
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T
C:Species: Lactococcus lactis phage BK5-T
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C:Accession: T13256
R:Boyce, J.D.; Davidson, B.E.; Hillier, A.J.
Appl. Environ. Microbiol. 61, 4089-4098, 1995
A:Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and
A:Reference number: Z17646; MUID:96064422
A:Accession: T13256
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1904 <BO>
A:Cross-references: EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579.1

Query Match 4.5%; Score 182; DB 2; Length 1904;
Best Local Similarity 20.7%; Pred. No. 0.012;
Matches 189; Conservative 83; Mismatches 260; Indels 380; Gaps 55;

QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPG-----ILXVRTDIGMYRWDANGR 50
Db 838 AKPSDYTWS--LIRNGDKDGTGKDGKGVAGKGVGKIKTTVITYALSSG---TDKPNTG 891
QY 51 W---IPLLDVWGNWNGYGVVSIADPINTNKVAAVGYMYSWDPNDGAILRSSDQGA 107
Db 892 WTSQVPTL-----VKGOYLTKTWVT---YTDG-----SSETG- 921
QY 108 TWOTLTPFLKLGMPGRGMRGLAVDPNNDNILYFCAPSGKGLWRSTDGATW-SQMTN 166
Db 922 -YSVTYI--AKDGNNGNDIAGDKGVGKIKTTVYAVGTSG-----TTAPASGWNQSVPN 973
QY 167 FP-----DVGTYIANPTDTTGIQVQSDIQQVVMVAFDKSSSLGQASKTIFVGVADPN 218
Db 974 VPAGQFLWTKTVWYTDN-TSETGYSVAMVMGVKDGKDPGNNGTN-----GIA---- 1020
QY 219 PVFWSRDG-----GATWQAVPG----APTG----FIP--HKGVF-----DP 249

Db 1021 ----GKDGKIGKATAIYQASPNGTTAPTCTWSASVPPVAKGSFLWTRTITWYTDNTTET 1076
QY 250 VNHVLIATISNTGCGPDGSGGVKESVTSCTWTRISVPSTANDYFGYGLITDRQH 309
Db 1077 GYAVAYMGTNGNG-HDGFPG-----KDGTKIKTTIT-----YAGTSGTTP 1118
QY 310 PNTIMVATQISWPPDITIFRSTDGGATWTR-IWDMTSYPNRSLRYVLDISAEPWLTFGVQ 368
Db 1119 PN-----NGMTSTVP-TVAEGNLYLTKTVWYTD--NTS----- 1149
QY 369 PNPPVPSPKLGWDEAMAI-----DPFNSDRMLYGTGATLYATNDLTKWDGGQIHAPM 423
Db 1150 -----ETGYSVAMVMGVKDGKDPGNN-----GTNGIAGKDG----- 1180
QY 424 VKLETAVNDLISP-PSGAPLISALGDGGFTHADVATVP--STIFTSPVFT----- 473
Db 1181 -KGKATAIYQASPNGTTAPT-----GTWSASVPPVAKGSFLWTRTITWYTDNTTE 1231
QY 474 TGTSDVYAEIN-----PSIIVRAGSEDPSSOPNDRHVAFASTDGKNWF 516
Db 1232 TGAVAYMGTNGNNGHDGPGKDGTKIKTTITYAGTSGTTPPNN-----GW- 1279
QY 517 QGSEPGGVTTGGTVAASADG---SRFVW-----APGDPGPVYVAVG 555
Db 1280 -----TSVPPTVAEGNLYLTKTVWYTDNTSETGYSVAMVMGVKDGKDP----- 1323
QY 556 FGNSWAASQGVPAQAQIRSDRVNPKTFYALNSGTFYRSTDGGVTFQVPAAG----- 606
Db 1324 -GNN--GTNGI-AGKDGKIGKATAIYQASPNCTTAPTGTWSASVPPVAKGSFLWTRTIW 1379
QY 607 -----LPSSGAVGVN-----FHAVPGKEG-----LWLAASSGLYHSTNGSS-- 644
Db 1380 TYTDNTETGYAVAYMGTNGNNGHDGPGKDGTKIKTTITYAGTSGTTPPNNGTSTV 1439
QY 645 -----W-----SAITGVSSA-VNVGFKSAPGSSYPVAVFVGTIGGVGTG-- 682
Db 1440 PTVAEGNLYLTKTVWYTDNTSETGYSVAMVMGVKDGKDPGNN-----GT-NGIAGK 1491
QY 683 -----AYRSDDCGTTWVLNDQHOYGNWGOAI---TGDHANLRV----- 720
Db 1492 GKIKATAIYQASPNGTT-----APTGTWSASVPPVAKGSFLWTRTITWYTDNTTE 1543
QY 721 -----YIGTNG 726
Db 1544 TGAVAYMGTNG 1555

RESULT 10
F70846
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70846
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: F70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2523 <CO>
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAAL17115.1; PID:e125
C:Experimental source: strain H37RV
A:Gene: PPE

Query Match 4.3%; Score 175.5; DB 2; Length 2523;
Best Local Similarity 21.8%; Pred. No. 0.044;

RESULT 12
AE2254
hypothetical protein alr3588 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AE2254
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. (strain PCC 7120)
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2254
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-908 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA875287.1; PID:g17132721; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3588

Query Match 4.2%; Score 171; DB 2; Length 908;
Best Local Similarity 20.7%; Pred. No. 0.021;
Matches 161; Conservative 98; Mismatches 272; Indels 246; Gaps 41;

QY 71 IAADPINTNKVAAVGMVMTNSWDPNDGAILRSSDQG---ATWQITPLPFFKLGGMPCGRM 127
Db 131 LSAEIVQONNI-----NDNTIANARNIGTFNTTITPTVDYVSGS----- 171

QY 128 GERLAVDPNDNLIYFGAPSGKGLMRSTDSGATWSQMTNPDPVGTYIANPTDTGTGYSODI 187
Db 172 GQ--VIDQND---YYS-----FTLTNSGTV-----DINLSLNGTDT--LYADL 208

QY 188 QGVVWVAFDKSSSLGQASKTIFGVADPNPVFWSRGGATWQAVP---GAPTGRIPKH 244
Db 209 QLI-----SSNSVIQTSATV-----GTSLESISRSLAAGTYIRAY 245

QY 245 GVFPDPVNHVLYI-----ATSNLTGGPYD-----GSSGDMVKFVSFV 278
Db 246 SQSDPGNYLNFNSADPPDAGGNTSDTSTPINLPATFSEIISDOVSLGSSDYQFTLA 305

QY 279 SGWTRISPVPSDTDANDYFGYGLTIDRQHPNTIMVATQISWPDITIFRSTG----- 333
Db 306 SASLVEIQFTSLTADANLY-----LQTONGNILSSQPGTALDAVRLSLNAGTYNIL 358

QY 334 ---GATWTRIDWTSY-----PNRSL-----RVVLDISAEPWLTFGV 367
Db 359 VNRGSTETAQYTLISGFAQAIGNQAPNSTTLALNLSNPISLNEFVGNIDTNDYKFTV 418

QY 368 QPNPPVPSPKLGWMD-----BAMADP--FNSDRMLYGTGATLYATNDLTWKDS----- 414
Db 419 NGTTEIN-----LDLSILSYLLDPQLVADVOILNSGGTQVAISNQTGNSNESINTIL 472

QY 415 -GGQIHAPVWKGLEETAVNDLISPSPGAPLISALGDLGGFTHADYVAVPSTIFTSPVFT 473
Db 473 GAGTYFIRVYTSGLANTFYDLNLTASQALLVQDINPTGNSDPANLTGLNTLY-----FT 528

QY 474 TGTSDVDAEL---NPSIIVR---AGFDPSSQ---PNDRHVAFSTD--GGKNW--FOG--- 518
Db 529 ANDGINGVQLWSNGDITTLRLKSNISFNPNLIVFNRLYFAASNDTFFGRELWEYNGTV 588

QY 519 -----SEPGGVTTGGT-----VAASADGSRFVWA-PGDPGPQPVVYAVGFNGSWA 561
Db 589 NRISDINVAGNSNPGNLTAVAGNKLFFTAVDNDSIRKLWYNGTNYNLVDVNASFNSST 648

QY 562 ASQGVFANAQIRSDRVNPKTFYALSNCTFYRSTG--GVTFQFVAACLPSSGAVGVNFWHA 619
Db 649 PTTFTTFNNQL-----FFTAQNNSQLWSTDGTGGTQVVISAGGITHKSTPRNL--T 696

QY 620 VPKEGDLWLAASGLYHSTNGSSSAITGVSSA---VNVFGKSAPESSYPVAVFVVG 675
Db 697 VVG--NTLYFTANNG-----TSGHEIWOYQNGTGTASLLEDITPGNNSFAP--SSLTAV---- 745

QY 676 TIGGVTGAVRSDDCGTTWVLINDDOHYG--NW---GOAITGDHANLRRVYIGTNGR 727
Db 746 -----GNTLIFVTDSDNDFNLELWKSGDTAAGTD-----IIGTDGQ 781

RESULT 13
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11208551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 4.2%; Score 170.5; DB 2; Length 5188;
Best Local Similarity 20.3%; Pred. No. 0.22;
Matches 191; Conservative 82; Mismatches 269; Indels 401; Gaps 47;

QY 27 GAPGILYVFTDIGMYRWDAAANGRWIPLLDWGWNNMGYVYSIAADPINTNKVAAVG 86
Db 4083 GANEFAQIISTDNCA-----TWVN-----VTVAADSLN----- 4109

QY 87 MYTNSWDPNDGAILRSSDQGATWQITPLPFLGGMNPGMGRGLAVDPNN----- 137
Db 4110 -----WSYVDGRTLNGT--TTWQVRV--DLAGNV-GATSSQSALIDTVPNAQVITIAS 4159

QY 138 -----DNILYF-----GAPSGKGLNRSTDSGATWSQMTNFPDVGTYIAN 176
Db 4160 ISDTGSSATDFTISDTMLTGLSLGAGLASGEVAQISLDSGATWTLTNGTQWYITDS 4219

QY 177 PTDTTG---YQSDIQGVVWVAFDKSSSLGQASKTIFGVADPN---NPVFWSRDGGATW 230
Db 4220 RLTLDGSYVQVR-----LDLAGNTGPVSKTVVVDINPTATPTIVSYTDDVG--- 4269

QY 231 QAVPGAPTGFPHKGVFDPVNHVLYIATSNLTGCPYDG-----SSGDV----- 272
Db 4270 -----QRQGTLSLQ-----ATDDTPLLNGVLSAPLASGEVYVLYRNGLLGA 4313

QY 273 -----WKFS---VTSQWT---RISVPVSTDTANDYFGYGLTIDRQHPNTIMVAT- 317
Db 4314 VTMVGALNWTYSDSLVSGAYTYSARVVDLAGNITSSDF---VLIVDTSIPTTLAQITS 4370

QY 318 -----OISW-----PDT--- 325
Db 4371 QTRDRTPIISGVITAALASQYVEVINGKTYTSEPGGAVVVDPAHNTWYVOLPOTDAL 4430

QY 326 -----IIFRSDGGATWTRIDWTSYPNRSLRYVL 355
Db 4431 TVSATAYTVAQVKSSAGNNGNANISNGTVVNAADYDTPWTTASKTTAW---GLTYGL 4487

QY 356 DISAEPWLTFGVOPNPVPSPKLGWDEAMADPENSDRMLYGTGATLYATNDLTWKDSG 415
Db 4488 D-SHGMWTVLANQVVMQSTDP-LTWSKTALT-----LYOSGNN-YATSIADYDRN 4535

QY 416 GQIHAPVWKGLEETAVNDLISPSPGAPLIS---ALGDLGGFT-HADYVAVPSTIFTSP 470
Db 4536 G-----TGDLEFTRDDYGTGYINGFTNNGDGT-----FSSA 4566

QY 471 V-FTTGTSDVDAELNLSIIVRAGSFDPSOPNDRHVAFT--DGGKNWFGSGEPG----- 522
Db 471 V-FTTGTSDVDAELNLSIIVRAGSFDPSOPNDRHVAFT--DGGKNWFGSGEPG----- 522


```
Db 4567 IQVTGTLTWYGSi-----VAFDKEGDYLDWFWDAGGPDsNT 4605
QY 523 -----GVTTGGTVAASADGSRFVWAPDGPQPVVAVGEGNSWAASOGVPANAQIRSDR 576
Db 4606 FLNNAGTLVGNSTTSNGSGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4655
QY 577 V-----NPKTFYALS-----NGTFYRSTGGVTGFQPVAAAGLPSGGAGVGMFHAVPGK-EG 625
Db 4656 VQHTYLNLYTLSSLLINGNGTFVWGQNTNTFLSGAGSGAMSSSVMTWADFDDGDM 4715
QY 626 DLWLAASSG-----LYHSTNG----- 641
Db 4716 DLFLPASQGRANYSGLLENTINGVLCPCVAVGATATTYASQFSLAVDWNHDLMDIARIAQ 4775
QY 642 -GSSWAITGVSSAVNVGKSPAGSSYPAVFVVGIGVGT-----GAYR-----SDDCG 690
Db 4776 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4827
QY 691 TTWVLNDQHOYGNNGQAITGDHANLRVYIGTNGRGIVYGD 733
Db 4828 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4860

RESULT 14
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAV>
A;Cross-references: GB:BA000007; PIDN:BA033965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain R1MD 0509952
C;Genetics:
A;Gene: ECs0542

Query Match 4.2%; Score 168.5; DB 2; Length 5291;
Best Local Similarity 20.1%; Pred. No. 0.3;
Matches 190; Conservative 82; Mismatches 270; Indels 401; Gaps 46;

QY 27 GAFGILYVRTDIGMYRWDAAWRWIPLLDWGNNWNGVYVSIADPINTNKVWAAVG 86
Db 4186 GANEFAQISTDNGA-----TWVN-----VTVAADSLN----- 4212

QY 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGGNMPGRGMGERLAVDPNN----- 137
Db 4213 -----WSYVDGRLTNGT--TTWQVRV--DLAGNV-GATSSQSALIDTVPNAQVLTIAS 4262

QY 138 -----DNILYF-----GAPSGKLMRSTDSGATSQMTNFPDVGVIYAN 176
Db 4263 ISDTGSSATDFITSDMTLTLTSLGSLAGLASGEVAQISLDSGATWTLTNGTQWYTDs 4322

QY 177 PTDTTG---YQSDIQGVWVAFDKSSSLGQASKTIFVGVADPN---NPFWSRDGGATW 230
Db 4323 RLTDGSIYVQVRV-----LDLAGNTGPVSKTVVVDVINPTATPTIVSYTDDVGQRQ 4375

QY 231 QAVPGAPTGFIPKHGVDPNVHLVYIATSNTPGPGYDG-----SSGDV----- 272
Db 4376 GTLSSSQ-----ATDDTTPLLNGVLSAPLASGEVYLYRNGLLGA 4416

QY 273 -----WKFS---VTSGWT---RISPVSTDDTANDYFGYSLGIDIDRHPNTIMVAT- 317
Db 4417 VTMVGALNWTYSGLVSGAYTYSARVVDLAGNITSSDF---VLTVDTSIPTTLAQITS 4473

QY 318 -----QISWV---PDT--- 325
```

```
Db 4474 QITROTPPIISGVITRAALASGOYVEVWINGKTYTSEPGAVVVDPAAHNTWYVQLPDTDAL 4533
QY 326 -----IIFRSTDGGATWTRIWDTWTSYPNRSRYVL 355
Db 4534 TVSATAYTVAQVKSSAGNANNISNGTVTNAAIDYPTPTWTTASKTTAW---GLTYGL 4590
QY 356 DISAEPWLTFGQPNPVPSPKLGWMDAEMADPFPNSDRMLYGTGATLVATNDLTWKDSG 415
Db 4591 D-SHGMMTVLANQQVMQSTDP-LTWSKTALT-----LYQSGNN-YATISSIADYDRN 4638
QY 416 GQIHAPMVKGLEETA VNDLISPPSGAPLIS-----ALGDLGGFT-HADVTAVPSTIFTSP 470
Db 4639 G-----TGDLFITRDDYCTGYINGFTNNGDGT-----FSSA 4669
QY 471 V-FTTGTSDVYAEALNPSPVIRAGSFPDPSOPNDRHVAFT--DGGKNWFQGSPEP----- 522
Db 4670 IQVTGTLTWYGSi-----VAFDKEGDYLDWFWDAGGPDsNT 4708
QY 523 -----GVTTGGTVAASADGSRFVWAPDGPQPVVAVGEGNSWAASOGVPANAQIRSDR 576
Db 4709 FLNNAGTLVGNSTTSNGSGSATVGG-----AVTGYLSLNEGSGVDLNNDRIDL 4758
QY 577 V-----NPKTFYALS-----NGTFYRSTGGVTGFQPVAAAGLPSGGAGVGMFHAVPGK-EG 625
Db 4759 VQHTYLNLYTLSSLLINGNGTFVWGQNTNTFLSGAGSGAMSSSVMTWADFDDGDM 4818
QY 626 DLWLAASSG-----LYHSTNG----- 641
Db 4819 DLFLPASQGRANYSGLLENTINGVLCPCVAVGATATTYASQFSLAVDWNHDLMDIARIAQ 4878
QY 642 -GSSWAITGVSSAVNVGKSPAGSSYPAVFVVGIGVGT-----GAYR-----SDDCG 690
Db 4879 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQS 4930
QY 691 TTWVLNDQHOYGNNGQAITGDHANLRVYIGTNGRGIVYGD 733
Db 4931 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4963

RESULT 15
A28452
proteoglycan core protein precursor, cartilage - rat
N;Alternate names: aggrecan
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 21-Jan-2000
C;Accession: A92623; A28453; A28454; A28455
R;Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 262, 17757-17767, 1987
A;Title: Complete primary structure of the rat cartilage proteoglycan core protein de
A;Reference number: A92623; MUID:88087070
A;Accession: A92623
A;Molecule type: mRNA
A;Residues: 1-2124 <DOE>
R;Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 263, 10040a, 1988
A;Reference number: A30069
A;Contents: annotation; revision to residue 698
R;Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 261, 8108-8111, 1986
A;Title: Partial cDNA sequence encoding a globular domain at the C terminus of the ra
A;Reference number: A23835; MUID:86250698
A;Accession: A23835
A;Molecule type: mRNA
A;Residues: 1856-2124 <DO2>
A;Cross-references: GB:M13518; NID:g206104; PIDN:AAAA1836.1; PID:g206105
R;Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 262, 17768-17778, 1987
A;Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-t
A;Reference number: A28453; MUID:88087071
A;Accession: A28453
A;Molecule type: protein
A;Residues: 20-37, 'W', 39-60, 'E', 62-64, 'X', 66-69; 70-83; 84, 89-148, 'L', 150-238, 'S', 240, '

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:53 ; Search time 28.05 Seconds
(without alignments)
1021.478 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	168	4.2	2124	1 PGCA_RAT	P07897 rattus norv
2	167.5	4.2	1119	1 AL53_CANAL	O74623 candida alb
3	166	4.1	1468	1 N153_RAT	P49791 rattus norv
4	157	3.9	809	1 QUITA_ACICA	Q59086 acinetobact
5	153	3.8	1150	1 APNU_PIG	P12021 sus scrofa
6	152	3.8	1196	1 AMTB_PAEPO	P21543 paenibacill
7	149	3.7	2132	1 PGCA_MOUSE	Q61282 mus musculu
8	148.5	3.7	776	1 ISOA_PSEAY	P10342 pseudomonas
9	148.5	3.7	776	1 ISOA_PSESP	P25927 salmonella
10	148.5	3.7	1953	1 BIGA_SALTY	P26501 pseudomonas
11	146	3.6	2109	1 PGCA_CHICK	P07898 gallus gall
12	145.5	3.6	790	1 QUITA_XANCU	Q9xg78 xanthomonas
13	145	3.6	342	1 Y48L_SNY3	P73069 synechocyst
14	145	3.6	747	1 GUNB_CELFI	P50400 cellulomona
15	143.5	3.6	937	1 NUN8_RAT	P49793 rattus norv
16	142.5	3.5	995	1 AGAA_VIB57	P48839 vibrio sp.
17	142	3.5	647	1 NANH_MICVI	Q02834 micromonop
18	142	3.5	757	1 DHET_GLUOX	O05542 gluconobact
19	142	3.5	760	1 YBIL_ECOLI	P75780 escherichia
20	139.5	3.5	1045	1 GUNB_CELFI	P26225 cellulomona
21	139.5	3.5	2333	1 PGCA_CANFA	Q28343 canis famli
22	139	3.4	3063	1 CAIC_HUMAN	Q99715 homo sapien
23	138	3.4	827	1 XANP_XANS2	Q60106 xanthomonas
24	137.5	3.4	872	1 GUXA_CELFI	P50401 cellulomona
25	137.5	3.4	1103	1 VG37_BPAR1	Q99065 bacterioph
26	137.5	3.4	2415	1 PGCA_HUMAN	P16112 homo sapien
27	136.5	3.4	522	1 NU62_HUMAN	P37198 homo sapien
28	136.5	3.4	781	1 NANH_VIBCH	P37060 vibrio chol
29	136	3.4	2628	1 HAGA_PORGI	Q51845 porphyromon
30	134.5	3.3	2344	1 POLN_RHDV	P27410 rabbit hemo
31	134	3.3	890	1 BCN5_CLOPE	P08696 clostridium
32	134	3.3	1014	1 NANH_CLOSE	P29767 clostridium
33	133.5	3.3	966	1 FIB1_PETMA	P02674 petromyzon

ALIGNMENTS

RESULT 1

ID	PGCA_RAT	STANDARD;	PRT;	2124 AA.
AC	P07897;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).			
DE	Protein) (CSPCP).			
GN	AGC1 OR AGC.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88087070; PubMed=3693370;			
RA	Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;			
RT	"Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones.";			
RL	J. Biol. Chem. 262:17757-17767(1987).			
RN	[2]			
RP	REVISION TO 698.			
RA	Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;			
RN	J. Biol. Chem. 263:10040-10040(1988).			
RL	[3]			
RP	SEQUENCE OF 1856-2124 FROM N.A.			
RX	MEDLINE=86250898; PubMed=2424893;			
RA	Doegge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;			
RT	"Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan.";			
RL	J. Biol. Chem. 261:8108-8111(1986).			
CC	-!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.			
CC	-!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).			
CC	-!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.			
CC	-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.			
CC	-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 4 LINK DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.			
CC	-!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -			

34	133	3.3	920	1 NU98_HUMAN	P52948 homo sapien
35	132	3.3	886	1 VGP3_EBVA8	Q07284 epstein-bar
36	132	3.3	1433	1 SUBF_BACSU	P16397 bacillus su
37	131.5	3.3	1026	1 VG37_BPT4	P03744 bacterioph
38	131.5	3.3	1199	1 N121_RAT	P25291 rattus norv
39	131.5	3.3	2090	1 N214_HUMAN	P35658 homo sapien
40	131	3.2	532	1 SPG7_DICDI	P22698 dictyosteli
41	131	3.2	1090	1 GUXB_CELFI	P50899 cellulomona
42	130.5	3.2	1341	1 VG37_BPT2	P07067 bacterioph
43	130.5	3.2	1365	1 GTFS_STRDO	P29336 streptococ
44	130	3.2	1025	1 SLAP_CAUCR	P35828 caulobacter
45	129	3.2	656	1 THIG_SVNY3	Q55710 synechocyst

AC	074523;	
AD	15-JUL-1999 (Rel. 38, Created)	
AE	15-JUL-1999 (Rel. 38, Last sequence update)	
AF	15-JUL-1999 (Rel. 38, Last annotation update)	
AG	Agglutinin-like protein 3 precursor.	
AL	ALS3	
AO	Candida albicans (Yeast).	
AP	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
AQ	Saccharomycetales; mitosporic Saccharomycetales; Candida.	
AR	NCBI_TaxID=5476;	
AS	[1]	
AT	SEQUENCE FROM N.A.	
AV	STRAIN=I161;	
AW	MEDLINE=98309840; PubMed=9644209;	
AX	Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;	
AY	"Candida albicans ALS3 and insights into the nature of the ALS gene	
AZ	family.";	
BA	Curr. Genet. 33:451-459(1998).	
BB	!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.	
BC	!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).	
BD	-----	
BE	This SWISS-PROT entry is copyright. It is produced through a collaboration	
BF	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
BG	the European Bioinformatics Institute. There are no restrictions on its	
BH	use by non-profit institutions as long as its content is in no way	
BI	modified and this statement is not removed. Usage by and for commercial	
BJ	entities requires a license agreement (See http://www.isb-sib.ch/announcement/	
BK	or send an email to license@isb-sib.ch).	
BL	-----	
BM	EMBL; U87956; AAC39486.1; -	
BN	Cell adhesion; Glycoprotein; Repeat; Signal.	
BO	SIGNAL	
BP	1 17	
BQ	FT CHAIN 18 1119 AGGLUTININ-LIKE PROTEIN 3.	
BR	FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.	
BS	FT REPEAT 433 468 1-1.	
BT	FT REPEAT 469 504 1-2.	
BU	FT REPEAT 505 540 1-3.	
BV	FT REPEAT 541 576 1-4.	
BW	FT REPEAT 577 612 1-5.	
BX	FT REPEAT 613 648 1-6.	
BY	FT REPEAT 649 684 1-7.	
BZ	FT REPEAT 685 720 1-8.	
CA	FT REPEAT 721 756 1-9.	
CB	FT REPEAT 757 792 1-10.	
CC	FT DOMAIN 399 404 POLY-THR.	
CD	FT DOMAIN 450 455 POLY-THR.	
CE	FT DOMAIN 557 563 POLY-THR.	
CF	FT DOMAIN 593 597 POLY-THR.	
CG	FT DOMAIN 630 635 POLY-THR.	
CH	FT DOMAIN 666 671 POLY-THR.	
CI	FT DOMAIN 702 707 POLY-THR.	
CJ	FT DOMAIN 738 743 POLY-THR.	
CK	FT DOMAIN 774 777 POLY-THR.	
CL	FT DOMAIN 1044 1047 POLY-THR.	
CM	FT CARBOHYD 471 471 N-LINKED (GLCNAC. .) (POTENTIAL).	
CN	FT CARBOHYD 543 543 N-LINKED (GLCNAC. .) (POTENTIAL).	
CO	FT CARBOHYD 579 579 N-LINKED (GLCNAC. .) (POTENTIAL).	
CP	FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).	
CQ	FT CARBOHYD 687 687 N-LINKED (GLCNAC. .) (POTENTIAL).	
CR	FT CARBOHYD 723 723 N-LINKED (GLCNAC. .) (POTENTIAL).	
CS	FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).	
CT	FT CARBOHYD 845 845 N-LINKED (GLCNAC. .) (POTENTIAL).	
CU	FT CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).	
CV	FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. .) (POTENTIAL).	
CW	FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. .) (POTENTIAL).	
CX	FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. .) (POTENTIAL).	
CY	SEQUENCE 1119 AA; 119927 MW; 6A3FB3FC8C879A71 CRC64;	
DQ		
EQ		
FQ		
GQ		
HQ		
IQ		
JQ		
KQ		
LQ		
MQ		
NQ		
OQ		
PQ		
QQ		
RQ		
SQ		

[illegible]

Query Match 4.2%; Score 167.5; DB 1; Length 1119;
Best Local Similarity 18.7%; Pred. No. 0.022;
Matches 170; Conservative 98; Mismatches 340; Indels 303; Gaps 42;
QY 7 TWSNVAIGGGGFVDGIVFNAGAGILYVRFDIG-----MYRWDAANGRW--IPL-LDWV 58

GN NUP153.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 602-613; 622-645 AND 971-993.
 RC STRAIN=BUFFALO; TISSUE=Liver;
 RX MEDLINE=93137325; PubMed=8422679;
 RA Sukegawa J., Blobel G.;
 RT A nuclear pore complex protein that contains zinc finger motifs,
 RT binds DNA, and faces the nucleoplasm.;
 RL Cell 72:29-38(1993).
 CC -!- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
 CC COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
 CC ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
 CC TERMINAL RING STRUCTURE OF THE NUCLEOLAR CAGE.
 CC -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
 CC -!- PTM: PHOSPHORYLATED IN INTERPHASE, HYPERPHOSPHORYLATED DURING
 CC MITOSIS. MAY PLAY A ROLE IN THE REVERSIBLE DISASSEMBLY OF THE
 CC NUCLEAR PORE COMPLEX DURING MITOSIS.
 CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
 CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
 CC NUP1, NUP1, POM 121 AND MAMMALIAN P62.
 CC -!- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L06821; ; NOT_ANNOTATED_CDS.
 DR InterPro: IPR001876; Znf-RanBP.
 DR Pfam: PF00641; Znf-RanBP; 4.
 DR SMART: SM00547; Znf-RBZ; 4.
 DR PROSITE: PS01358; ZF_RANBP2_1; 4.
 DR PROSITE: PS01359; ZF_RANBP2_2; 4.
 KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding;
 KW Phosphorylation.
 FT DOMAIN 4 15 GLY-RICH.
 FT DOMAIN 442 446 POLY-GLY.
 FT ZN_FING 657 687 RANBP2-TYPE 1.
 FT ZN_FING 721 750 RANBP2-TYPE 2.
 FT ZN_FING 789 818 RANBP2-TYPE 3.
 FT ZN_FING 846 875 RANBP2-TYPE 4.
 SQ SEQUENCE 1468 AA; 152823 MW; C3DFD9697C556A7C CRC64;

 Query Match 4.1%; Score 166; DB 1; Length 1468;
 Best Local Similarity 21.8%; Pred. No. 0.036;
 Matches 144; Conservative 71; Mismatches 259; Indels 186; Gaps 31;

 QY 106 GATWQITPLP--FKLGNMPPGRNGERLAVDPNNNDILYFGAGSGKGLWRSTDSGATWSQ 163
 DB 884 GTSSLLNPAPSAFKFGIPSSSSGLSTQTTSTGN---FKEGDGGKFLGTSSDSGSTNTM 939
 QY 164 MTNF-----PD-----VGTYIANPTDTFTGYQSDIQGVV 191
 DB 940 NTNFKPKPTGDKFGVLPLDPSKPEIKNDKNDNFQFGPSSGLSNPASSAPFQ---GVS 996
 QY 192 WVAFDKSSSSLGQASKTIF---GVADPNPNPFWSRDGGATWQAVGAPTGFPHKGVED 248
 DB 997 TLGQOQKEELPOSSAGSFAGVANPSSAAI---DVTTSNKSGFNFTDITKSV-- 1051
 QY 249 PVNHVLIATNTGGPGVGGSGDGVKWFVSTGTWTRISVPVPSDTDANDYFGYSLTIDR- 307
 DB 1052 -----SVTPFYKTTKTEAKEDASATKGFT-FGKVDSSAALSPSMFVLGRTEEKQ 1100
 QY 308 QHPNTIMVATQISWPPDTIIF-RSTD-----GGATWTRIDWNTSYPNRSLRYV 354

DB 1101 QEPVT-----STSLVFGKKADNEPKCPQVFSFGNSEQTKD----- 1136
 QY 355 LDISAPFWLTFGQVPPPPVSPKLGWDEAMADPNPNSDRMLYGTGATLYATNDLTKWDS 414
 DB 1137 -ESSRPTFTFSVAK-----PSVK-----ESDLAKATFAFGQTNTTT--DQ 1176
 QY 415 GGGIHTAPMVKGLEETAVNDLISPPSGAPLISALGDL-GGFTHADVTAVSTIF---TSP 470
 DB 1177 GA---AKPAFSLNSS-----SSSSTPATSSASIFGSSSTSSSPVAAFFVGOASNP 1227
 QY 471 VFTT--GTSVDYAEIIMPISIIRAGSDPSPQNDRHVA-----FSTDGKKNMFGSGEPGV 524
 DB 1228 VSSSAFGNSAESSTSQPLLPQDQK--PATTSASAAPPEFVGTGASSN---STVSSGF 1282
 QY 525 TTGCTGAASADGSRFVWAPGDCQPVVYAVG-----FGNSWAASQGVANAIQRSDRVN 578
 DB 1283 TFGATTTSSSSGSFVFGTGHSAFASAPAFGANTPTFGOSQASQ-----PN 1330
 QY 579 PKTFYALNSGTFTYRSDGGVTQFQVAAAGLPSSGAGVGMFHAVPGKGDLMWLAASSGLYHS 638
 DB 1331 PPSFGSISSSTALFSA-GSOPVPPPTFGTVSSSSQPPVFGQOPSQS-----AFSGGTANA 1384
 QY 639 TNGSGSWSALTGV-----SSAVNVGFGKSAPG-----SSYPAVFVVGTTIG 678
 DB 1385 SSVFQFGSSTTFNFTNNPNSGVFTFGASPSTPAAAAQPSSGGGFSQSPASFTVGSNG 1444

 RESULT 4
 QUITA_ACICA STANDARD; PRT; 809 AA.
 ID QUITA_ACICA
 AC Q59086;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Quinate/shikimate dehydrogenase [Pyroloquinoline-quinone]
 DE (EC 1.1.99.25) (NAD(P)-independent quinate dehydrogenase).
 GN QUITA.
 OS Acinetobacter calcoaceticus.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OC Acinetobacter
 OC NCBI_TaxID=471;
 RN [1]
 RP SEQUENCE FROM N.A.; AND CHARACTERIZATION.
 RC STRAIN=BD413 / ADP1;
 RX MEDLINE=95095936; PubMed=8002591;
 RA Elsemore D.A., Ornstom L.N.;
 RT "The pca-pob supraoperonic cluster of Acinetobacter calcoaceticus
 RT contains quiaA, the structural gene for quinate-shikimate
 RT dehydrogenase.";
 RL J. Bacteriol. 176:7659-7666(1994).
 RN [2]
 RP SEQUENCE OF 1-20 FROM N.A.
 RC STRAIN=BD413 / ADP1;
 RX MEDLINE=96011389; PubMed=75923351;
 RA Elsemore D.A., Ornstom L.N.;
 RT "Unusual ancestry of dehydratases associated with quinate catabolism
 RT in Acinetobacter calcoaceticus.";
 RL J. Bacteriol. 177:5971-5978(1995).
 CC -!- FUNCTION: CAN ACT EITHER ON QUINATE OR ON SHIKIMATE.
 CC -!- CATALYTIC ACTIVITY: Quinate + pyrroloquinoline-quinone = 5-
 CC dehydroquininate + reduced pyrroloquinoline-quinone.
 CC -!- CATALYTIC ACTIVITY: SHIKIMATE + PYRROLOQUINOLINE-QUINONE = 3-
 CC DEHYDROSHIKIMATE + REDUCED PYRROLOQUINOLINE-QUINONE.
 CC -!- COFACTOR: PQQ.
 CC -!- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; FIRST STEP. THIS PATHWAY
 CC ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO
 CC PROTOCATTECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE
 CC PATHWAY.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- INDUCTION: BY PROTOCATTECHUATE.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.
 CC -----

RP SEQUENCE OF 1-936 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=72;
RX MEDLINE=87165765; PubMed=2435707;
RA Kawazu T., Nakanishi Y., Uozumi N., Sasaki T., Yamagata H.,
RA Tsukagoshi N., Udaoka S.;
RT "Cloning and nucleotide sequence of the gene coding for enzymatically
active fragments of the Bacillus polymyxa beta-amylase.";
RL J. Bacteriol. 169:1564-1570(1987).
RN [2]
RP SEQUENCE OF 689-1196 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=72;
RX MEDLINE=89123046; PubMed=2464578;
RA Uozumi N., Sakurai K., Sasaki T., Takekawa S., Yamagata H.,
RA Tsukagoshi N., Udaoka S.;
RT "A single gene directs synthesis of a precursor protein with beta-
amylase activities in Bacillus polymyxa.";
RL J. Bacteriol. 171:375-382(1989).
RN [3]
RP SEQUENCE OF 1-776 FROM N.A.
RC STRAIN=ATCC 8523;
RX MEDLINE=87231094; PubMed=2438660;
RA Rhodes C., Strasser J., Friedberg F.;
RT "Sequence of an active fragment of B. polymyxa beta amylase.";
RL Nucleic Acids Res. 15:3934-3934(1987).
RN [4]
RP DISULFIDE BOND, AND MUTAGENESIS OF CYSTEINE RESIDUES:
RX MEDLINE=91215008; PubMed=1827035;
RA Uozumi N., Matsuda T., Tsukagoshi N., Udaoka S.;
RT "Structural and functional roles of cysteine residues of Bacillus
polymyxa beta-amylase.";
RL Biochemistry 30:4594-4599(1991).
CC -1- FUNCTION: THE PRECURSOR PROTEIN IS PROTEOLITICALLY CLEAVED TO
PRODUCE MULTIFORM BETA-AMYLASES AND A 48 kDa ALPHA-AMYLASE AFTER
SECRETION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
polysaccharides so as to remove successive maltose units from the
non-reducing ends of the chains.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO FAMILY 14 OF
GLYCOSYL HYDROLASES (BETA-AMYLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 13 OF
GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15817; AAA85446.1; -;
DR EMBL; Y00150; CAA68344.1; -;
DR PIR; A29130;
DR PIR; A29108; A29108.
DR PIR; A32251; A32251.
DR HSP; P36924; I892.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR001554; Glyco_hydro_14.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase.C; 1.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Multifunctional enzyme; Hydrolyase; Glycosidase; Signal;
KW Polysaccharide degradation; Repeat.
FT SIGNAL 1 35
FT CHAIN 36 1196 BETA/ALPHA-AMYLASE.
FT DOMAIN 36 454 BETA-AMYLASE.
FT REPEAT 455 558

FT REPEAT 565 668
FT DOMAIN 669 1196 ALPHA-AMYLASE.
FT DISULFID 118 126
FT ACT_SITE 124 124
FT ACT_SITE 198 198
FT MUTAGEN 118 118 BY SIMILARITY.
FT MUTAGEN 126 126 C->S: 5 FOLD DECREASE IN ACTIVITY.
FT MUTAGEN 358 358 C->V: 20 FOLD DECREASE IN ACTIVITY.
FT MUTAGEN 358 358 C->S: 60 FOLD DECREASE IN ACTIVITY.
FT CONFLICT 1 1 M -> MGL (IN REF. 3).
FT CONFLICT 67 67 N -> S (IN REF. 3).
FT CONFLICT 100 100 N -> D (IN REF. 3).
FT CONFLICT 154 154 S -> N (IN REF. 3).
FT CONFLICT 177 177 E -> Q (IN REF. 3).
FT CONFLICT 227 228 NA -> KS (IN REF. 3).
FT CONFLICT 330 330 G -> S (IN REF. 3).
FT CONFLICT 425 425 N -> S (IN REF. 3).
FT CONFLICT 493 493 D -> A (IN REF. 3).
FT CONFLICT 532 532 S -> L (IN REF. 3).
FT CONFLICT 559 559 A -> T (IN REF. 3).
FT CONFLICT 665 665 A -> T (IN REF. 3).
FT CONFLICT 681 681 D -> N (IN REF. 3).
FT CONFLICT 686 686 T -> A (IN REF. 3).
FT CONFLICT 725 728 APTS -> VFSP (IN REF. 3).
FT CONFLICT 736 736 N -> K (IN REF. 3).
FT CONFLICT 741 741 N -> S (IN REF. 3).
FT CONFLICT 758 758 S -> N (IN REF. 3).
SQ SEQUENCE 1196 AA; 130893 MW; A41EA6B70F257064 CRC64;
Query Match 3.8%; Score 152; DB 1; Length 1196;
Best Local Similarity 21.1%; Pred. No. 0.18;
Matches 167; Conservative 72; Mismatches 282; Indels 270; Gaps 45;
QY 61 NNWG-----YNGVVSIAAD-----PINTNKV-WAAVGMVYNSWDNDGAILRS 102
Db 53 NDWGSFKQLQTLKNNGVYAITTDVWVGYSVESAGDNDFMSYKYKYANA-----V 102
QY 103 SDGATWQITPLPKLGGNMPGRGMLAVDPNNDNLYFGAPSGKGLWRSTDGATWS 162
Db 103 KEAGLKWPIIITHKCGN-----VGD-----DCN-----IPLPS-----WLSSKGSDEM 143
QY 163 QMTNFPDVGTYIANPTDTTGY-QSDIQGVYVAVFDKSSSLGQASKTIFVGVADPNPNVF 221
Db 144 QF-----KDESYANSEALSPWSGTGKQYDELYASFAENFAGYSIIPKIY 190
QY 222 WSRDGGATWQAVGAPGTFT-----PHKGVF-----DPVNHVLYIATSYTGGP 264
Db 191 LS-----GGPSGELRYPSPYPAAGWSYFGRGKFAQYTETAKNAFRTAMNDKYGS 239
QY 265 YDSSGDVWKFVSYTSYSGTWTRISVPSTDTANDYFGYSLTIDRQHPNTIMVATQISWPD 324
Db 240 LDKINA-AWGTKLTS--LSQINPTDGDGYTNGGY-----NSAYGKDFLSWQS 286
QY 325 TI-----IPRSTDG-----GATWTR-----IWDWTSYPNR 349
Db 287 VLEKHLGVIGAAAHKNFDSVFGVRIGAKISLGHWMNNPAMPHGTEQAGGYDY----NR 342
QY 350 SLRYVLDISAEPWLT-----GVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGA 401
Db 343 LIQKFKDADILD--LTFTCLEMSDSGTAPNYSLPS-----TLVDTVSS--IANAKGV 389
QY 402 TLVATNDLTKWDSGGQIHAPMVK---GLEETAVNDLISPPSGAPLISALGDLDGFF--- 454
Db 390 RLNGENALPTGGSGFQKIEKITKFGYHGFTLLRNNLVN-NDGSP-----TGELSGFKQY 444
QY 455 -----THADVTAVPSTFTSPVF-----TTGTSVDYAEIN--PSIIV 489
Db 445 IISKAPDNNGGTGNKVTIYKGFNSPYIHYRPAAGSWTAAPGVKMQDAEISYAKITV 504
QY 490 RAGSFDPSQPNDRHVAFTDGGKNWFQGS-----EPGGVTGTGTVAAASDG 536
Db 505 DIGS-----ASQLEAAFN-DGNNNDNSNTKNYSFSTGTSTYTPGNSGNAGTITSGAPA 557

QY 537 SRFWAPGDPG-----QPVVYAVGF-----GNSWAASOGVPAN-----AQIRS 574
 Db 558 GA---NPGDGGTNTKVVYVYKGFNSPYTHYRPAAGGSWTAAPGVKMQDAEISGYAKITV 614
 QY 575 DRYNPXTFYAL-----SNCT-FVRSDDGGVTQPVAAAGLPSSCAVCVMFHAVPGKE 624
 Db 615 DIGASQLEAFAFNDGNNNDNNKNTKNTFTSTGYTTP-----GNGAAGTITRTGAP--S 667
 QY 625 GDWLAAASSGLY-----HSTNGGSSWSAIVTGVSSAVNVGFGKSAPGSSYPFAVFGVIGGV 680
 Db 668 GSV-LSVVTSTYATDLNEVTGPIOTEKLSGV--LNVSTSTYAPNSN--GVEVTAQTEAP 722
 QY 681 TGAYRSDDCGT 691
 Db 723 SGAFTSMDLGT 733

RESULT 7
 PGCA_MOUSE STANDARD; PRT; 2132 AA.

AC 061282; Q64021;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN AGC1 OR AGC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Cartilage;
 RX MEDLINE=95104847; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fueleop C., Horvath P., Doege K.J., Glant T.T.;
 RT "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan.";
 RL Genomics 22:364-371(1994).
 RN [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.;
 RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene.";
 RL Nat. Genet. 7:154-157(1994).
 CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION IN THE G1 DOMAIN).
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L07049; AAC37670.1; -
 DR EMBL; S73722; AAB32160.1; -
 DR EMBL; S73721; AAB32160.1; JOINED.
 DR HSSP; P98066; LTSG.
 DR MGD; MGI:99602; AGC.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_V.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXXSG.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR InterPro; IPR001304; lectin_C.
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF02339; SGXXSG; 62.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00406; IGV; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS01241; LINK; 4.
 DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
 KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 GI-A.
 FT DOMAIN 152 247 GI-B.
 FT DOMAIN 253 349 GI-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 587 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.
 FT DOMAIN 1917 2132 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 580 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1922 1933 BY SIMILARITY.
 FT DISULFID 1950 2042 BY SIMILARITY.
 FT DISULFID 2018 2034 BY SIMILARITY.
 FT DISULFID 2049 2092 BY SIMILARITY.
 FT DISULFID 2078 2105 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CONFLICT	386	386	L -> V (IN REF. 1).	
FT	CONFLICT	413	416	GAYT -> AVH (IN REF. 1).	
FT	CONFLICT	454	489	SGIDLFAEYAGFGNSYQLGGFPQGSWENGLFRDS -> TYVICLNLGPSAAATRTSWDSSRRVRVVEVSPRQ (IN REF. 1).	
FT	CONFLICT	555	556	WP -> S (IN REF. 1).	
FT	CONFLICT	650	657	AFRKAHPA -> RSARHIP (IN REF. 1).	
SEQ	SEQUENCE	776 AA;	83626 MW;	F738BF8040246169 CRC64;	
Query Match 3.7%; Score 148.5; DB 1; Length 776;					
Best Local Similarity 20.6%; Pred. No. 0.17;					
Matches 164; Conservative 90; Mismatches 330; Indels 213; Gaps					
QY	27	GAGCILVVRTDIGMYRWDAANGRWIPLLDVGNWNNWNGVYSIAADPTINTNKVAAVY 86			
DB	96	GITGAVY-----YCYRAWGNPVPYASNWKGQSAGFVSDVDANGDREPNKLLDDY 147			
QY	87	MYNSWDPNDCAILRSSDQATWQITPLPKLGGNPPGRMGGRERLAVDPNNNDILYFGAP 146			
DB	148	AOEVSQDP-----LNFSPQ-----NGNYFASCASYR-----TTDSGIY--AP 182			
QY	147	SGKGLWRSTDGATWSQMTNFPDV--GTYIANPTD--TTGVQSDIQGVVWVAFDKSS--SS 201			
DB	183	KGVVLPVSTOSTGKTPRAQKDVIYEVHVRGTFEODTSIPAQYRTTYGAGLKASYLAS 242			
QY	202	LGOASKTIFVGAD-----PN---NPVFW-----SRDGATWQAVPGAPTG-F 240			
DB	243	LG-VTAEFLPVQETONDANDVVPNSDANQNYGMYTENYFSPDRYAYNKAAGGTAEF 301			
QY	241	IPHKGVDPVNVHLYIATSNWTGPGYDGGSCDVWKFVSTGVTWTRISPVST----- 291			
DB	302	QAMVQAEPHNAIKYMDV-----VYNHTAEGGTWSSOPTTATYSWRGLDN 348			
QY	292	-----DTANDYF-GYSGLTIDRQHPNTI---NVATQISWMPDITIFRSTDGGATWTRIW 341			
DB	349	ATYYELTSGNOYFVDNTGICANFNTYVAQNILVDSLAYWANTM---GVDGFR----- 400			
QY	342	DWTS-YPNRSRLRVLDISAPPLWTFGVQPNPVPSPKLGWMDPAMAID-----PFNSDRML 396			
DB	401	DLASVLGNSCLNGAYTASA-----PNCPNGYFNDAADSNVAINRILREFTVVRPAA 451			
QY	397	YGTGATLYATNDLTKWDSGGQIH-IAPMVKGLPE-----TAVNDLISPPSGAPLI 445			
DB	452	GGSLDLFA---EPWAIGNSYQLGGFPQGSWENGLFRDSLRAQNEL-----GSMIT 502			
QY	446	SALGDLGGFFHADVTAVPSTIFTSPVFTTGTSTVDYAEALNPISIVR-----AGS 493			
DB	503	YVTDANDFSGS-----SNLFOSGRSPWNSINFIDVHDGMTLKDVIYSCNGANNNSQAMP 556			
QY	494	FDPSSOPNDRHVAFTSDGGK---NNFQSGEPGVTTGGTV---AASADGSRFVWAPGDP 546			
DB	557	YGPS-----DGGTSTNYSDWQMSAG---TGAADVORRAARTGMAFEML--SA 599			
QY	547	GQPVVYAVGFGNSWAASQGPANAAQIRSDRVNPKTF-YALNSNGTFYRSTDGGVTFQPVAA 605			
DB	600	GTPLMQG---GDEYLRTLQCNNNAYNLDSSANWLTYISWTTDQSNFYTAQRLLIAFRKAHP 656			
QY	606	GL-PSSGAVGVMEHAPGREGDLWLAASSGLYHSTNG---GSSWSAIT--GVSSAVNVVG 658			
DB	657	ALRPSS-----WYSGSLTWTPQSGAVADSNYWNNTSNYAIAYAIN-- 697			
QY	659	FGKSPGSSYPVYVVGITGGVTCAYRSDDCGTWTWLINDDQHQYGNWQAATITGDHANLR 718			
DB	698	GFSLGDSNIIYVAYNGWSSSVFTLLPAPPSGTQWYRVYDT---CDWNDGASTFVAPGS 752			
QY	719	RVYIGNRGIVYGDIG 735			
DB	753	ETLLIG--GAGTTYQCQC 767			

RESULT 9
ISOA_PSESP


```
Db 302 QAMVQAFHNAIKYMDV-----VYNHTAEGTWTSSDPTTATIIYSWGLDN 348
QY 292 -----DTANDYF-GYSGLFIDRQHNTI---MVATQISWMPDTIIFRSDGGATWRIW 341
Db 349 TTYVELTSGNOYFDNTGIGANTENTYNAQNLIVDSLAYWANTM---GVDGFR-----F 400
QY 342 DWTSP-YPNRSLRVLDISAEPWLTFGVQPNPPVPSKLGHWDEMAID-----PFNSDRML 396
Db 401 DLASVLGNSCLNGAYTASA-----PNCPNGYNFDAADSNVAINRLREFTVRPAA 451
QY 397 YGATGATLAFNDLTKWDGSGQIHF-IAPMVKGLLE-----TAVNDLISPPSGAPLI 445
Db 452 GGSGLDLFA---EPWAIGNSYQLGGFPGQSEWNGCLFRDSLRAQNEL-----GSMTI 502
QY 446 SALGDGGLGFFHADVTAVPSTFTSPTTGTSTVDYAEIPLSIIVR-----AGS 493
Db 503 YVTQDANDESGS-----SNLFOSGSRSPWNSINFDVHDGMDLTKDVISCGANNSSQAWP 556
QY 494 FDPSSQPNDRHVAFTDGGK---NMFOGSEPGVTTGGTV---AASADGSRFVWAPGDP 546
Db 557 YGFS-----DGGSTNYSDQGMASG---TGAADVORRAARTGMAFEML---SA 599
QY 547 GQPVVAVGFCNSWAASQGVPAANAQIRSDRVNPKTF-YALNSGTFFVRSDDGGVTFQPVAA 605
Db 600 GTPLMQG---GDEYLRFLQCNNAYNLDSSANWLTYSWTTDQSNFYTFQORLIATFRKAHP 656
QY 606 GL-PSSGAVGMFHAVPKEGDLWLAASSGLYHSTNG-----GSSWSAIT---GVSSAVNVG 658
Db 657 ALRPSS-----WYSGSLTWYQPSGAVADSNVWNTSNYATAYAIN-- 697
QY 659 FGKSAPGSSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDQHQYGNWGAITGDHANLR 718
Db 698 -GPSLGDNSIYYAYNGWSSSTFTLPPAPSGTQWRYVTD-----CDWMDGASTFVAPGS 752
QY 719 RVVIGTNGRGIVYGDIG 735
Db 753 ETLIG--GAGTTYGQCG 767

RESULT 10
BIGA_SALTY
ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=14028;
RC "Salmonella typhimurium rhs homolog.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT Lr2.v;
RL Nature 413:852-856(2001).
[3]
RN SEQUENCE OF 1-765 FROM N.A.
```

```
RC STRAIN=LT2;
RX MEDLINE=911100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting siroheme
RT cofactor.";
RL J. Bacteriol. 173:325-333(1991).
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF133696; AAD39458.1; -
DR EMBL; AE008859; AAL22340.1; -
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; bigA.
KW Virulence; Repeat; Signal; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1953 PUTATIVE SURFACE-EXPOSED VIRULENCE
FT DOMAIN 101 252 15 X 11 AA TANDEM REPEATS.
FT REPEAT 101 103 1 (INCOMPLETE).
FT REPEAT 104 113 2 (INCOMPLETE).
FT REPEAT 114 122 3 (INCOMPLETE).
FT REPEAT 123 133 4.
FT REPEAT 134 144 5.
FT REPEAT 145 155 6.
FT REPEAT 156 166 7.
FT REPEAT 167 177 8.
FT REPEAT 178 188 9.
FT REPEAT 189 199 10.
FT REPEAT 200 210 11.
FT REPEAT 211 221 12.
FT REPEAT 222 232 13.
FT REPEAT 233 243 14.
FT REPEAT 244 252 15 (INCOMPLETE).
FT CONFLICT 207 207 D -> DRGDDVTPPD (IN REF. 1).
FT CONFLICT 514 514 A -> R (IN REF. 3).
FT CONFLICT 1698 1698 D -> N (IN REF. 1).
FT CONFLICT 1795 1798 OYLE -> ITLO (IN REF. 1).
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 3.7%; Score 148.5; DB 1; Length 1953;
Best Local Similarity 20.5%; Pred. No. 0.51;
Matches 158; Conservative 83; Mismatches 274; Indels 257; Gaps 34;

QY 44 WDAANG-----RWIPLLDVGWNNNGVGVSTAADPINTKNVAAVGYMTNSWD---P 94
Db 1344 WNTATGVINGINPDAPLINLGRYNGFNQGTINVGQD---NAV--AISGGTSSYVNLV 1397
QY 95 NDGAILRSSDQAGTWTQITPLPFKLGGNMPG-----RGMGERLAVDPNDNLIYFCAPSGKG 150
Db 1398 NSGTINVTGEQK-----EDGTNGTGLIGKGNATNTINTADGVINVIYADSYA 1448
QY 151 LWRSTDSGATWSQMTNPPDVGTVIANPTDTGYQSDIQGVVWVAFDKSSSLGQASKTIF 210
Db 1449 FGKTRAIINNGEINLLCDSGCDIYAP-GTGTQNDUNGADIVIPDATTAPTEGS--IP 1505
QY 211 VGVADPNPVFWSRDRGGATQWQVPGAPTGFPHKGVFDPVNHVLYIATSTNGPYDSSG 270
Db 1506 TTPADPNAPQOLSN-----YIVGTNA---DGSSG 1531
```



```
QY 271 DVKFSV-----TSGTWTRISPPVSTDTANDYFYCSGLTIDRQHPNTIMVATQ 318
D 1532 TLKANLVIGNVKVDGFTSGADTTVVVDNAFTGSNIQADNIT-----STS 1580
QY 319 ISWPPDIIIFRSTDDGGATWTRINDWTSYPNRSLRYVLDISAEPMLTFTGVQPNPVPSPKL 378
D 1581 VVW--NAQGSODADGNY-----DVTMTKNAYADVATDSSVS----- 1614
QY 379 GWDEMAAIDPFNSDRMLY-----GTGATLYATNDLTWKDSSGQ----- 417
D 1615 ---DVAQALDAGYTNNELYTSNLVGTITAEI---NSALKQVSGQAQTTVFREARVLSNRF 1668
QY 418 --THIAPMVK--GLEETAANDLISPPSGAPLISALGDIGFTHADVTAVPSTIETSPVFTT 474
D 1669 MLADAAFPQIKDGLAFNVV-----AKGDP---RAELGNDTQYDMLALROTLDLTA--SQ 1716
QY 475 GTSVDYAEALNPSIIVRAGSDPSPQPNDRHVAFSTDGKNWFQSGPGVTTGGTVAASA 534
D 1717 NLITLEY-----GIARLDG 1729
QY 535 DGSRFVWAPDGPVYVAVGFG--NSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYR 592
D 1730 DGSK---TAGDNGLTGYSOFFGLKHSMAFDEGLAWNLSRYDVHNLDSRSVAYGDNVK 1786
QY 593 STD-----GGVTFQPVAAGLPSSGAVGVNF--HAVPG--KE---GDLWLAASS 633
D 1787 IADSDMRQOYLEFRSEGAFTTMMGDALKVTPYAGVKFRHTMEDGYKERSAGDFNLMSNS 1846
QY 634 -----GLXHSITGSSWSAITGVSSAVNVGFKS-----APGSSYPAPFVVG 675
D 1847 GNETAVDSIVGLKLDYAGKGSWSATATLEGPNLSYSKSORTASLOGAAGQSF--GVDDGQ 1905
QY 676 TIGVTC-----AYRSDDCGTTWLVINDHQYGVNMGQAITGDHANLRVY 721
D 1906 KGGVGNLATIGYKYSND-----TALHLDAYQWKEDGISDKGFMLNVKTF 1952

RESULT 11
PGCA_CHICK STANDARD; PRT: 2109 AA.
ID AC P07898; Q90991; Q90820; Q91047; Q90810;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
DE Gallus.
GN AC1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEHORN; TISSUE=Embryo;
RX MEDLINE=94043149; PubMed=8226878;
RA Li H., Schwartz N.B., Vertel B.M.;
RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia."
RT J. Biol. Chem. 268:23504-23511(1993).
RN [2]
RP SEQUENCE OF 1042-1559 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90307744; PubMed=1694853;
RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II. Nucleotide sequence of cDNA clone and localization of the S103L epitope."
RT J. Biol. Chem. 265:12088-12097(1990).
RN [3]
RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC TISSUE=Cartilage;
```

```
RX MEDLINE=93111968; PubMed=1339285;
RA Chandrasekaran L., Tanzer M.L.;
RT "Molecular cloning of chicken aggrecan. Structural analyses."
RL Biochem. J. 288:903-910(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=94107258; PubMed=8280087;
RA Chandrasekaran L., Tanzer M.L.;
RL Biochem. J. 296:885-887(1993).
RN [5]
RP SEQUENCE OF 1492-1610 FROM N.A.
RC STRAIN=WHITE LEHORN; TISSUE=Chondrocytes;
RX MEDLINE=95128519; PubMed=7827752;
RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
RT "Molecular basis of nanomelia, a heritable chondrodystrophy of chicken."
RL Matrix Biol. 14:297-305(1994).
RN [6]
RP SEQUENCE OF 1894-2109 FROM N.A.
RX MEDLINE=89008500; PubMed=3170613;
RA Tanaka T., Har-El R., Tanzer M.L.;
RT "Partial structure of the gene for chicken cartilage proteoglycan core protein."
RL J. Biol. Chem. 263:15831-15835(1988).
RN [7]
RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
RX MEDLINE=86259736; PubMed=3450082;
RA Sai S., Tanaka T., Kosher R.A., Tanzer M.L.;
RT "Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein."
RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS ACTIVELY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES
CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT (CHONDRODYSTROPHY) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS. AGGECAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGECAN/VERSIAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L21913; AAB19128.1; -
DR EMBL; M38187; AAA48731.1; -
DR EMBL; M88101; -; NOT_ANNOTATED_CDS.
DR EMBL; S74657; AAC60751.1; -
DR EMBL; S74656; AAC60751.1; JOINED.
DR EMBL; J04028; AAA48719.1; -
```



```
RESULT 12
QIUA_XANCJ STANDARD; PRT; 790 AA.
AC QXAD78;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable quinolate dehydrogenase [Pyroloquinoline-quinone]
DE (EC 1.1.99.25).
GN QIUA.
OS Xanthomonas campestris (pv. juglandis).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=44291;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C5;
RX MEDLINE=20063481; PubMed=10594704;
RA Lee Y.-A., Lo Y.-C., Yu P.-P.;
RT "A gene involved in quinolate metabolism is specific to one DNA homology
RT group of Xanthomonas campestris.";
RL J. Appl. Microbiol. 87:649-658(1999).
CC -1- CATALYTIC ACTIVITY: Quinate + pyroloquinoline-quinone = 5-
CC dehydroquininate + reduced pyroloquinoline-quinone.
CC -1- COFACTOR: PQQ (BY SIMILARITY).
CC -1- PATHWAY: QUINIC ACID CATABOLIC PATHWAY; FIRST STEP. THIS PATHWAY
CC ALLOWS GROWTH OF BACTERIA WITH QUINATE BY ITS CONVERSION TO
CC PROTOCATECHUATE AND SUBSEQUENT METABOLISM BY THE BETA-KETOADIPATE
CC PATHWAY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL PQQ DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF109471; AAD38453.1; -.
DR InterPro: IPR001479; Bac_PQQ.
DR InterPro: IPR002372; Bac_PQQ_repeat.
DR Pfam: PF01011; Bacterial_PQQ; 6.
DR PROSITE: PS00363; BACTERIAL_PQQ.1; FALSE NEG.
DR PROSITE: PS00364; BACTERIAL_PQQ.2; FALSE NEG.
KW Oxidoreductase; PQQ; Quinate metabolism; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 77 94 POTENTIAL.
FT TRANSMEM 106 126 POTENTIAL.
SQ SEQUENCE 790 AA; 82896 MW; B75F29B52A49F6F6 CRC64;

Query Match 3.6%; Score 145.5; DB 1; Length 790;
Best Local Similarity 21.6%; Pred. No. 0.26;
Matches 166; Conservative 87; Mismatches 239; Indels 275; Gaps 50;

QY 60 WNNWGY--NGVVSIAADPIITNKVAAVGYMTNSWDPNDGAILRSDQGATWQITLPLPK 117
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 151 WSAYGNITDGSREFAALDQINRSGRPAAG---SPGPTTPEGIANSDNGAEDQLTLQ-- 205
QY 118 LGGNMPGRGRLAVDPNNDNLTLYGAPSKGLWRSTDGATWSQNTNPDVGTY--- 173
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 206 -----VGEKVFICTPHNNLIADASTGKQLWR-REINATSSVWQRCRGLGYDADA 255
QY 174 -----IANPTDTTG-----YQSDIQG-VVWV-----AFDKSSSSILGQASKT 208
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 256 ALPAPSVAVNSPIAAVTAQAGNCRRRLFTNTIDGLRIADATGAFCCQFGSGNGQVDLK 315
QY 209 IFVGVA--DP-----NNPVFSRBD--GGATWQAVPGAPT---GFIPHKGVDPDNNHVLIA 257
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
DB 316 AGLGAAPDPFYQLTSPPLVAGTTVVGGRT--RADDNQVTDMPGGVVRGSMSPVRS---A 370
QY 258 TSNTGGPYD-----GSS-----GDVM-KFSVTSGTWTRISPV--PSTD-----TANDY 297
DB 371 GLDPGNPHDRQAPAAAGSSVYRSTPNVWAPMSYDAAMNTVFLPLGGFSTDLYGAERTALDH 430
QY 298 -FGYSLGITIDROHPNTIMVATQISWMPDTIIFRSTDGGATW-----TRIWDWTSYPNRS 350
DB 431 RYCASVALD-----ATTGAEKWVYQTVHNDLWDF----- 460
QY 351 LRVLDISAEPWL--TFGVQVNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDL 409
DB 461 -----DLPMQPSLIDFPNQDGSHTPAVVG----- 485
QY 410 TKWDSGGQIHIAPMVKGLEETAVNDLISPPSGA-----PLISALGDLG--GFTH 456
DB 486 TK---AGQIYVLDRAATGKPLTEVREV--PVKGSIDIAHEQYAPTQPLSVGMPQIGTKHLTE 540
QY 457 ADV---TAVPEST-----IFTSVFTTGTSDVYAEINPSIIVRAGSFDPSSQP 500
DB 541 SDMWGATAMDQMLCRITAFKQMYEGLYTAB---GTDVSL-----SF-PGSL- 582
QY 501 NDRHVAFTSDGGKNWFGGSEPGVTTGGT--VAASADGSRFVWA---PGD----- 545
DB 583 -----GGMNW-----GGLSTDVPVHDVVFANDMRGLWQVMI PADTRKAEAGGG 626
QY 546 -----PGQPVVYAVGFGNSWAASQGVPAQAQIRS--DRVNPKT----FYALSNGT 590
DB 627 EAVNTGMGAVPLKGTAYAVN-KNRFLSALGIPCAQAPPYGTLSDIKLTRISIAHQVPGTV 685
QY 591 YRSTDGGVTQF--PVAAGLPSSGAV-----GVMFHVPGKEGDLWLAASSGLYHSTNGSS 644
DB 686 QDTPGPGIKMHLPIGMPTLGGTLSTQGGLVF--TAGTQ--DYVLA-----FDSATGKEL 738
QY 645 WSAITGVSSAVN--VGKKSAPGSSYPAVFVVGITGGVGTGAYRSDDCG 690
DB 739 WKGRLPVGSOGGPITYVSHKTKQY-----VVISAG---GARQSPDRG 778

RESULT 13
Y48L_SYNV3 STANDARD; PRT; 342 AA.
AC P73069;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ycf48-like protein.
GN SLR2034.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT 'Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.';
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE YCF48 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```


Db 585 EV--YRGSTLIVARTGSHVTGLSAAATAYTFTVRAVDAAGNVS-AASAPVGVTAPDPT 641

QY 665 -GSSYPVAVVGTIGGVTGAYRSDCCGTT----WVL 695

Db 642 TGSCAVTYTANGSGGFTAAVTLTNGTTALSGWTL 677

RESULT 15

NU98_RAT

ID NU98_RAT STANDARD; PRT; 937 AA.

AC P49793;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Nuclear pore complex protein Nup98 (Nucleoporin Nup98) (98 kDa nucleoporin).

DE nucleoporin).

GN NUP98.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Liver;

RX MEDLINE=95254630; PubMed=7736573;

RA Radu A., Moore M.S., Blobel G.;

RT "The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex.";

RL Cell 81:215-222(1995).

CC -!- FUNCTION: PLAYS A ROLE IN THE BIDIRECTIONAL TRANSPORT ACROSS THE NUCLEOPORIN COMPLEX (NPC). THE REPEAT DOMAIN HAS A DIRECT ROLE IN THE TRANSPORT.

CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.

CC -!- DOMAIN: CONTAINS G-L-F-G REPEATS.

CC -!- PTM: THE N-TERMINUS IS BLOCKED.

CC -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L39991; AAC42054.1; -

DR InterPro: IPR004325; Nucleoporin_FG.

DR Pfam: PF03093; Nucleoporin_FG; 26.

KW Nuclear protein; Transport; Repeat.

FT DOMAIN 889 893 POLY-GLU.

SQ SEQUENCE 937 AA; 97809 MW; 6B1F108C7390307A CRC64;

Query Match 3.6%; Score 143.5; DB 1; Length 937;

Best Local Similarity 20.4%; Pred. No. 0.41;

Matches 137; Conservative 78; Mismatches 222; Indels 233; Gaps 34;

QY 143 FGAP-SGKGLWRSTDGATWSQMTNPDVGTI--ANPTDTTGYQSDIQGVVWVAFDK-- 197

Db 6 FGTPFGGTGGGTGFT---STFGQNTGF---GTTSGGAFGTSAFGSSNTGGLFGNSQTKPG 59

QY 198 ---SSSLGQ-----ASKTIFVGVDAPNNPNVFWSRDGGATWQAVPGAPT 238

Db 60 GLFGTSSLSQPAVSTSTGFCGFTSTGTSNSLF-GTANTGTSLSFSSQNNNA---FAQNKPT 114

QY 239 GFTHPKGVF--DPNVHVLVYATNTGNGPYDGGSGDVWVKFS-----VTSGTWTRISPVPTSD 292

Db 115 GF-----GNFCTSSSGGLFGTNTTNPFCNTSCSLFGPSSFTAAPTATIKENPPTGTD 170

QY 293 TANDYFCYSGLTIDRQHPNTIMVATOISWPPDIIIFRSTDGGATWTRINDWTSYPNRSR 352

Db 171 T-----MVKAGV-----STNISTKHOCITAMKEYSKSLE 200

Search completed: July 2, 2002, 09:18:03

Job time: 371 sec

QY 353 YVLDISAEPW-----LTFGVQPNPPVPSPKLGMMDEAMADPFN--SDRML 396

Db 201 ---ELRLEDYQANKRGPNQVAGAGTTTGLFGSSPATSSATGLFSSSTTNSAFSYGQNKTA 257

QY 397 YGTGATLYATNDLTWKWDSGGQIHIAPVWGLEETAYNDLISPPSGAPLISALDGLGGFTH 456

Db 258 FGSTTTGFGTN-----PGG-----LFGQQNQQTSLFSKPFG----- 289

QY 457 ADVTAVPSTIFTSPVTTTGTSDVYAEIPLNPSIIVRAGSFDSPSSOPNDRHVAFSTDGKNWF 516

Db 290 -QATTPNTGFS-----FNTSTLGPST-----NTWGLFGVT 321

QY 517 QGSEPGGVTTGTTVAASADGSRFVWAPGDPQVPVYAVGFGNSWAASQGVYANAIQRSDR 576

Db 322 QASQPGGLF--GTATNTSTGTAFGTGLFGQP---NTGFG---AVGSTLFGNNKLT--- 370

QY 577 VNPKTFFYALSNGTFYRSTDGGVTFQPVAAGLPSSGAVGVMFHAVPGKEGDLWLAASGL- 635

Db 371 -----TFGTSTTSAPSGFTTSGGL-----FGNKPRTLTLGTNTNTS 405

QY 636 ---YHSTNGGS---SWSAITGVSSAVNVGFGKSAFGSSYPAVF-----VVGTFIGVGTG 682

Db 406 NFGFGTNNSSSIFGSKPAAGTLGLTGLTGF--TALCAGQASLFGNNOPKIGPLG--TG 482

QY 683 AYRSDDCGTTWVLI-----NDDQHQYGNWQQAITGDHANLRRVYIGTNGRG 728

Db 463 AFCAPGFNTSTAILGFCAPQAPVALTDPNAAQA-----QAVLQQHLN-----S 506

QY 729 IVYGDIGGAP 738

Db 507 LTYSPEGDSP 516

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:22 ; Search time 91.42 Seconds

(without alignments)
1400.309 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTPQYTSNVAIGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2371	58.7	890	2	086727	086727 streptomyc
2	2052.5	51.1	996	2	09AQH0	09aqh0 caldicellul
3	2009	49.8	839	16	097KK0	097kk0 clostridium
4	1753.5	43.4	806	3	09P4T8	09p4t8 agaricus bi
5	1680	41.6	856	3	074170	074170 aspergillus
6	1160.5	28.8	707	16	09WYE1	09wye1 thermotoga
7	238	5.9	2468	16	0912M3	0912m3 pseudomonas
8	214	5.3	1800	2	09L948	09l948 pseudomonas
9	202.5	5.0	2174	16	092U08	092uu8 rhizobium m
10	189	4.7	3972	16	093139	093139 synechocyst
11	186.5	4.6	4199	16	P74440	P74440 synechocyst
12	185.5	4.6	818	6	09N1P0	09n1p0 bos taurus
13	185.5	4.6	997	2	09Z4I1	09z4i1 bacillus sp
14	184.5	4.6	998	17	09HLQ9	09hlq9 thermoplasma
15	182.5	4.5	848	2	09RK65	09rk65 streptomyc
16	182	4.5	1904	9	Q38319	Q38319 lactococcus

17	175.5	4.3	2523	16	053393	053393 mycobacteri
18	174.5	4.3	3930	16	09BE20	09be20 rhizobium l
19	167	4.1	1804	1	09P9A9	09p9a9 uncultured
20	167	4.1	3659	16	098LN6	098ln6 rhizobium l
21	166.5	4.1	902	16	09AAJ5	09aaJ5 caulobacter
22	166.5	4.1	3145	16	098MG7	098mg7 rhizobium l
23	166	4.1	699	16	007219	007219 mycobacteri
24	166	4.1	1203	5	09N5K0	09n5k0 caenorhabdi
25	164.5	4.1	695	2	0934G0	0934g0 pseudomonas
26	164	4.1	631	6	09N1P1	09n1p1 bos taurus
27	163.5	4.1	1047	3	093865	093865 candida alb
28	163.5	4.1	13055	5	009165	009165 caenorhabdi
29	163	4.0	1684	2	003658	003658 unidentified
30	160.5	4.0	3716	16	050379	050379 mycobacteri
31	159.5	4.0	1915	2	09REL0	09rpl0 acetivibrio
32	158.5	3.9	762	2	09RM63	09rm63 myroides od
33	158.5	3.9	1441	16	09CFL1	09cfl1 lactococcus
34	158	3.9	2232	5	091365	091365 caenorhabdi
35	156.5	3.9	695	2	09F9U2	09f9u2 pseudomonas
36	156.5	3.9	2204	16	007231	007231 mycobacteri
37	156	3.9	2117	3	096U11	096ull neurospora
38	156	3.9	3157	16	050378	050378 mycobacteri
39	155	3.8	809	5	09NAN8	09nan8 caenorhabdi
40	155	3.8	4180	16	0915N6	0915n6 pseudomonas
41	154.5	3.8	3570	4	099552	099552 homo sapien
42	154	3.8	779	2	0938A2	0938a2 stigmatella
43	153	3.8	1055	16	09A9P9	09a9p9 caulobacter
44	153	3.8	1115	4	09HD43	09hd43 homo sapien
45	153	3.8	13288	6	018758	018758 sus scrofa

ALIGNMENTS

RESULT 1	ID	086727	PRELIMINARY;	PRT;	890 AA.
086727	AC	086727;	DT	01-NOV-1998	(TREMBLrel. 08, Created)
	DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)		
	DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
	DE	PUTATIVE SECRETED CELLULOSE.			
	GN	SC5C7.30C.			
	OS	Streptomyces coelicolor.			
	OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
	OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
	OX	NCBI_TaxID=1902;			
	RN	[1]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN=A3(2);			
	RA	Seeger K.J., Harris D.;			
	RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
	RN	[2]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN=A3(2);			
	RA	Parkhill J., Barrell B.G., Rajandream M.A.;			
	RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
	RN	[3]			
	RP	SEQUENCE FROM N.A.			
	RC	STRAIN=A3(2);			
	RA	Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,			
	RL	Medline=97000351; PubMed=8843436;			
	RN	"A set of ordered cosmid and a detailed genetic and physical map for			
	RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
	RM	Mol. Microbiol. 21:77-96(1996).			
	DR	EMBL; AL031515; CAA20642.1; -			
	DR	HSSP; P07986; IEXG.			
	DR	InterPro; IPR002860; BNR.			
	DR	InterPro; IPR001919; CBD_2.			
	DR	Pfam; PF02012; BNR; 10.			
	DR	Pfam; PF00553; CBD_2; 1.			
	DR	SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;			

RC	STRAIN=Tok7B.1;				
RX	MEDLINE=20171169; PubMed=10706665;				
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,				
RA	Bergquist P.L.;				
RT	*Multidomain and multifunctional glycosyl hydrolases from the extreme				
RT	thermophile Caldicellulosiruptor isolate Tok7B.1.;				
RL	Curr. Microbiol. 40:333-340(2000).				
DR	EMBL: AF078038; AAK06388.1; -				
DR	HSSP: Q06851; INBC.				
DR	InterPro: IPR002860; BNR.				
DR	InterPro: IPR001956; CBD_3.				
DR	InterPro: IPR002965; P-rich_extensn.				
DR	Pfam: PF02012; BNR; 9.				
DR	PRINTS: PR01217; PRICHEXTENS.				
DR	ProDom: PD001947; CBD_3; 1.				
KW	Hydrolase.				
FT	NON_TER	996	996		
SQ	SEQUENCE	996 AA;	108275 MW;	3C72B6ED22F3C614	CRC64;
Query Match					
Best Local Similarity		51.1%;	Score 2062.5;	DB 2;	Length 996;
Matches 387;		Conservative	96;	Mismatches	238;
Indels		15;	Gaps		
Qy	4	QPTWSNVAI--GGGFFVDGIVFNEGAGILYVRTDIGMYRWDAAANGRWIPLLDWVGNN	62		
Db	35	QPYVKNVKIEGGGFTIGVFNPKENLVVVRTDIGAYRSTGGNTWTLQMDWVSFDD	94		
Qy	63	WYNGVYSIAADPINTNKV-WAAVGMVYNSWDPNDSGAILRSSDQATWQITPLPFKLGNN	121		
Db	95	WNLGVESIAATDPVDPNXXVILACQGTYYNSWTDMNGAILRSTDEGDTFEITPLPFKLGNN	154		
Qy	122	MPGRGMEERLAVDPNNDILYFCAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDTT	181		
Db	155	MPXNRLGERLAIDPNNNRILYLGREGXGLWKSSEDYGVSMKKVTSFPNPGTYIEDPCNP	214		
Qy	182	GYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWSRDGGATWQVAPGAPTGF	241		
Db	215	DYLNHITGVWVWVDFPTSGRPGCKIYVGVADKTTSIYYTKDGGTQWALPGQPTGLL	274		
Qy	242	PHKGVFDPVNHVLIATSNRTGGPYDGSQGVWKFVSCTGTRISPVPSDTANDYFGVS	301		
Db	275	PQRAKLSS--DGMLYITYSNTQTPYNGDYGEVWRYNTKTGEWKNISPMQAQDT---YFGV	330		
Qy	302	GLTIDRQHPNTIMVATQISWMPDTIIFRSTGGATWTRIMDWTSYPNRSRLRYVLDISAE	361		
Db	331	GLAVDAQNPVWVVAALSSWMPDTYIWRSIDGGETWKCIWENGYPNRTLHYNMDSAAP	390		
Qy	362	WLTFG-VQPNPPVSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWSDGGQTHI	420		
Db	391	WLNFQXTBPTPEVSPKLGWVGTLTIDPFNSDXMLYGTGAXLYGCDLTLNWDKQGNITI	450		
Qy	421	APVKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVTTCTSDVY	480		
Db	451	KVKAIGIEETSQVQALISPPVGPLFSALGDIAGFRHEDLEKAPNWTYVQPNMGTITDIF	510		
Qy	481	AEINPSIIVRAGSFDPSSQPNDRHVAFTSDGKNWFOG-SEPGVTTGTTGVAASADGRF	539		
Db	511	AEINPNFVWVGVNDKQWNPNTNRICFSYDGGKSWFGQNTPEPQGTSEGGTVAADAASAV	570		
Qy	540	VWAPGDPQPVVYAVFGFNSWAAASQGVYVAPNAQIRSDRVNPKTFYALSNGTYRSTDGGV	599		
Db	571	VWAP--KGAKVCYSTDNGKNWKECANVPSEAIYVSDRVNPNKTFYAFKNGKFYISADKGT	628		
Qy	600	F-OPVAAAGLPSGSAVGMFHAVPKGEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVG	658		
Db	629	FIESPAAGLPSG---NFKTVPGIEGDIWLGVNNGMWHSTDDGYSFVKISGVEDAASIG	684		
Qy	659	FGKSPAGSSYPVAVVVGTTIGVTGAYRSDCGTTWVLINDDQHOYGNWQQAITGDHANLR	718		
Db	685	FGKPAEGETIPATYTYAKINGVRGIFRSDDCDCTWIRINDDKHQFGCCANADITGDPVYG	744		
Qy	719	RVYGTNGRGIVYVDI	734		


```
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.; White O.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001712; AAC35393.1; -.
DR TIGR; TM0305; -.
DR InterPro; IPR002960; BNR.
DR Pfam; PF02012; BNR; 9.
SQ Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CDBA6A99F CRC64;

Query Match 28.8%; Score 1160.5; DB 16; Length 707;
Best Local Similarity 35.2%; Pred. No. 3.2e-57;
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

QY 6 YTSNVAIGGGGVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWGVNNGY 65
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 21 FEWKSVIEINGGFGVPGIIFHPASGLLYARTDVGLYRWDEETKRWKQLDFLRDQSDY 80
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 NGVYSTAADPINTNKVAAVGYMTNSWDNDGAILRSSDQATWQITPLP---FKLGGNM 122
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 MGVLVALDPSDPKRIYATMGYTDW-AGYGAILSEIDYGETWTIVNLDKYGKVGNE 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 PGRCMERLAVPNDNINILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTG 182
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 140 DGRNAGERLQVDFNFSVLFMGIT-KYGLWKSEDFGKNNKKVDSFST-----186
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 YQSDIOGVVWVAFDKSSSLGQASKTIFVGVADPNPNVFWSDGGATQAVPGAPGFTIP 242
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 -----SVTFVLFDKSGSGKSGPTPIFVGCSEPKG-IEFTEDGGTTWNVLNPLNDLIP 239
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 HKGVDPVNVHLYIATNSGTGGYDGGSSGVWKFVSTSGTWTRISPVSTDTANDYEGYS 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 LRQKIH--DGILVTLUSNALGNATRGAVMKYVIADQKWVDVTPMKGD-----FGYCG 291
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 LTIDRQHPNTIMVATQISWMPDTIIFRSTDGGATWTRIMDWTSYNRSRLYVLIDISAEPW 362
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 IDVOE--NVVIVSTLDRAWPHDEIFISLNGGETWRPLEKANF-----DINKAPW 339
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 363 LTFGVOPNPVPSKLGWDEAMADPFNSDRMLYGTATYATNLDLTKWDSGGQTHIAP 422
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 340 IK-----HWISD-VKIDPFDMNRAIFTGYGVWVTELYELKKSFEQ-----MGK 384
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 MYK-----GLRETAVDNLISPPSGAPLISALGDLGFTHADVTAVPSTFTSPVETTC 476
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 PVKWIFENKLEETVVLQVLPPIGERPLLSATADWGGFRHESLDTTPPSMY-KPLKWTSL 443
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 477 SVDYAELNPSITVRAGSPDPSSQPNDRHVAFTSDGKNNFQ-GSEPGVTTGG--TVAAS 533
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 444 GIAFAYONSKEVARVHTY--IYV---FLSYSDGGINNREIETVPEGITDGGRLSLAYS 497
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 534 ADCSRVFWAPDGPQPVYAVAGFNSWAASQGVPA---NAQTRSDRVNPKTFYAL---SNG 588
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 NDKGTLVWSPAN--HEVIVSSDKGSKWKKAISYPVPFNFYFPASDPVNPFSKFIEDWKG 555
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 589 TFRSTDDGGVTFOPVAAAGLPSSGAGVMPFA-----VPGKGDGLWLAAS-SGLYHSTNGGS 643
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 556 DFLISKDGKSPMK-CAKLPSPDNWVWVLSYFPVLAPDREGDILWLQWNLGRYSKDDGGI 614
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 644 SMSAITGVSAVNVGFKSAPSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDQHOY 703
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 615 TPERLGNVDIAYVIGFAPKPGTDXPIYLNGWVNGVYGIEMSTDGKTMRLNNDKHQF 674
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 704 GNWQAITGCHANLRVYIGTNGRIVGDI 734
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 675 G-WIHYMIGDMNEFGRIFLGTGGRGIIVEV 704
```

```
RESULT 7
Q912M3
ID Q912M3 PRELIMINARY; PRT; 2468 AA.
AC Q912M3;
DT 01-WAR-2001 (TREMREL. 16, Created)
DT 01-WAR-2001 (TREMREL. 16, Last sequence update)
DT 01-OCT-2001 (TREMREL. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA1874.
GN PA1874.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004613; AAG05263.1; -.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR003880; Phosphopant_attach.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00102; PHOSPHOPANTHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 5.9%; Score 238; DB 16; Length 2468;
Best Local Similarity 22.8%; Pred. No. 6.3e-05;
Matches 205; Conservative 88; Mismatches 314; Indels 292; Gaps 50;

QY 11 VAIGGGVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWGVNNGY-----65
Db 1084 VNLSGSSLSG---TAEPGSTVILTD-----GNGNPIAEVTADSGNWTTPSTP 1130
QY 66 --NG-VVSTAADPINTNKVAAVGYMTNS-----WDPNDGAILR-SSDQATWQITPLP 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1131 IANGTVVYVAQDASGNSPPATVTVDSAPPAPVLPNSGVVISGTAEGAGATVTLT---1187
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 FKLGGMPCGRGGERLAVDPNDNINILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIA 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1188 -DAGGNPIQGVTAQ-----GSGNWSFTP-----GTPLA 1214
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 N-----PTDTTGYQS---DIQGVVWVAFDKSSSL--GOASKTIFVGVADPN-N 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1215 NGTVIVATATDPTGNTGTPQAATTVDAVAPPAPVDPIDPSNGTTISGTAEAGAKVILTDGNG 1274
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 PV-FWSRDCGATWQAVPGAP-----TG-----FIPKGVDPVN 251
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1275 PIGETTADSGNSWSTPGTGPLANGTVVYNAVADPAGNTGPGSGTTVDAVAPNTPVVNP 1334
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 252 HVLYIATNTGGPY---DGSSGDVWKFVST-SGTW--TRISPVPS---TDTANDYFGVS 301
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1335 GNLLNCTAEPGSTVTLTDGNGNPIGTTADSGNSWSTFGSQLPNGTVVYVNTASDAAGNT 1394
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 GL-----TIDRQHPNTIMV-----ATQISWMPD---TIIFRSTDG-----ATWTRIND 342
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1395 SLPATTTVDSLSLPSIQVDPNSGVISGTADAGNTIIL--TDGNGNPIQGVTAADSGNWS 1452
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 WT-----STPNRSRLRYVLDISAEPWL--TFGVQPNPVPSPKLGWDEAMADPFNS 392
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1453 FTGPIPLDGTVVNVVARSNPVDSAPAVITVDGVAPAAV-----IDPSNG 1499
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 DRMLYGT---GATLYATNDLTKWDSGGQTHIAPVWKGLEETAVNDLISP-----PSGAPLI 445
```



```
Db 1500 TE-ISGTAAGATVILT-----DGGN-----PIQATADGSGNWTFTPTPLANGTVIN 1548
QY 446 SALGDLGGFTH-----ADTAVPSTIF--TSPVFTGTSDVYAEINPSIIVRAGSFDPS 497
Db 1549 AVAADPAGNTSGPASVTVDAIAPAPVNPNSGVISGT-----AEGATVILTDGNGNPI 1604
QY 498 SQPNDRHVAFTSDGGKNWF-----QSEPGGVF-----525
Db 1605 GQ-----VTADSGNWSFTPTCPLANGSVINALQAAGNNSPTSATVDSLAPAAPV 1657
QY 526 -----TCGTVAASADGRFVWAPDGPQPV--VTAVFGNSMAASQGVPAANAQIRSRVNP 579
Db 1658 IDPSNGSVIAGTAAGATVILTDGNGNPIQGVTAADSGN--WSFTPGTP-----1704
QY 580 KTFVALSNGTFYR-----STDGGV--TFQPVAAAGLP---SSGAVGVMFHVAAPGKEG 626
Db 1705 -----LSNGTVVNAVQAADAGANTSGPVSTTVDAVATPVIDPSNGVELSGTAEPGRVI 1759
QY 627 LWLAASSGLYHS--TNGSSWSAITGV-----SSAVNVGFGKSPAGSSYPVAVFVGTIGGV 681
Db 1760 LTDGNGNPIQTLADGSGNWSFTPGPLANGTVVNAVQAADPAGNTSGPASTTVDTVAPAT 1819
QY 682 GAYRSDCGTTWVLINDDQHQYGNWGOATGDHANLRVYIGTNGRGVYVDIGGAPSG 740
Db 1820 -----PVINPSN-----GSVITGT-AEVCAGKVLTDGNGNPIGETTADGSG 1859

RESULT 8
Q9L948 PRELIMINARY; PRT; 1800 AA.
AC Q9L948;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-BINDING OUTER MEMBRANE-LIKE PROTEIN (FRAGMENT).
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]_ID=303;
RP SEQUENCE FROM N.A.
RX MEDLINE=20225830; PubMed=10762233;
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RT "Genetic analysis of functions involved in adhesion of pseudomonas putida to seeds.";
RL J. Bacteriol. 182:2363-2369(2000).
RN [2]

SEQUENCE FROM N.A.
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF182518; AAF69021.1;
DR InterPro: IPR003644; Calx_beta.
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR InterPro: IPR002035; VWFA.
DR Pfam; PF00353; hemolysinCbind; 3.
DR PRINTS; PR00313; CABNDNGRPT.
DR SMART; SM00237; Calx_beta.1.
DR SMART; SM00327; VWFA.1.
DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 2.
DR PROSITE; PS50234; VWFA.1.
FT NON_TER 1
SQ SEQUENCE 1800 AA; 185196 MW; 409BFC017552F96E CRC64;
```

```
Query Match 5.3%; Score 214; DB 2; Length 1800;
Best Local Similarity 22.0%; Pred.No. 0.00093;
Matches 186; Conservative 107; Mismatches 319; Indels 232; Gaps 41;

QY 12 AIGGGGFDGIVNECAPGLLYVRTDIGMYRWDAAANGRWIPLLDVWGNWNCYGVVSI 71
Db 854 AAGNLOFFNGTAVNVAVGVVVSQADI-----TARNLKFVPALNOSGADNYGGNGVGNO 907
```

```
QY 72 AADPINTNKVMAAVGYTN--SWDPNDGAILRSSDQATWQITPLPKLGNMPPGRGMGER 130
Db 908 KAD-----YAQFKFPNDGTNL-----GSEVTMK 931
QY 131 LAVDPNNDN--ILYFGAP--SGKGL----WRSTDGATWSOMTFPPDVGYTIAN-----P 177
Db 932 VDPSPADKPTLSFGSADIESKGLTKREVMTSLKGLGTGGNGITGEDLKTVFANSANS 991
QY 178 TDTGTQSDIQGVVVAFAFDKSSSSLSGOAKT-----IFGVADPNPNPFWSRD 225
Db 992 SITTNVQSD-----GSVTAGTSGKTSGLIYLEAGKYVYTFGLADDS---FWVTI 1037
QY 226 GG-----ATWQAVQAGPTGFIPIHKGVFDPVNVHLYIATSNMTGGPYDGSDDVWKFVTS 280
Db 1038 GGGTIVTATWAGGGVSGTFTPTNTSGYPIE--VYHANQSGPGSYD-----LNIQVGS 1089
QY 281 TWTRISP-----VPSTDNDANDYFGYSGL--TIDRQHPNTIMVATOISWNPDTIIFRSTDG 333
Db 1090 AVTDLSSSNVYQNYTEMANAGLVSDLHTVNGQ-----SYDGYKLNKEGPEG 1138
QY 334 G-----ATWTRIMDWTSPNBSLR-----YVLDISAEPLWTFGVQPNPPVPSPKLGWM 381
Db 1139 GSVKLVGIETALDTDGSESLNVTLSGIPKGTVLSGAGHTVTVGT-----APVDVTGWK 1193
QY 382 DEAMADIP-----FNSDRMLYGT-----GATLYAT-----NDLTK 411
Db 1194 LSSLTLTPPAYKGSFDITVTSTATESLGSAITTCNIPVTYVYATYKASVGTSGNDTLT 1253
QY 412 WDSGGQIHIAPMVKGLEETAVNDLISPPSGAPLISALDGLGGTHADVTAVPSTIFTSPV 471
Db 1254 GSEGNDFIVAD-VSGLVNVQGNKYN-----NIAFMVDSGSGMSDKSIADAKTQLASVFNTLK 1308
QY 472 FTGTGS-----VDY-AELNPSIIVRAGSFDPSQPNDRHVAFTSDGKKNWFOGS-- 519
Db 1309 ASLGSDTSGTVNIFLVDFTQVKNKNAVLA--DPDALSCLKQVLNWSVGG--YGGGTN 1364
QY 520 -EPGGVTT-----GGTVAASADGSRFVWAPDGPQPVVYAVFGNSMAASQGVPAQAIRS 574
Db 1365 YEDAFKTTSNFFNSTWATSNKGAENLYFTTIDGKPYQO---SNESTNPISLWKNKSLD 1420
QY 575 DRVNPK-----TFYALSNGTFYRSTDDGGVTFOPVAAGLPSSCAGVMFHAVPCKEGDLW 628
Db 1421 DVVNVNRYKMGDTFSAWADATH-----KVEISSSGSVVVKVLTAT-TENRREGELV 1466
QY 629 LAASS--GLYHSTNGGS--SWSAITGVSSAVNVGFGKSPAGSSYPVAVFVGTIGGVTAIR 685
Db 1467 LSTKTVGTILHAQGDGYEFSLLDGTGYADYWNVYVSAAGST--ESFAV--LGGTNGLSK 1522
QY 686 SDDCGTTWVLINDDQHQYGNWGOAITG--DHANLRRYVIG-----TNGRGIV 730
Db 1523 VQAIGLNSDVTLNDLKPYSAGKPNIDPSDLAKAILGHSEATVPGADTIDSGNGNDII 1582
QY 731 YGDI 734
Db 1583 FGDL 1586

RESULT 9
Q92U08 PRELIMINARY; PRT; 2174 AA.
AC Q92U08;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL GLYCINE-RICH PROTEIN SMB21548.
GN SMB21548.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
```


RC	STRAIN=1021;
RX	MEDLINE=Z1936508; PubMed=11481431;
RA	Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Vorhoeher F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J., Golding B., Puehler A.;
RA	"The complete sequence of the 1,683-kb pSymB megaplasmid from the N2- fixing endosymbiont Sinorhizobium meliloti.";
RT	Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
RL	EMBL; AL603645; CAC49389.1;
KW	Plasmid; Hypothetical protein; Complete proteome.
DQ	SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;
SQ	

Query Match	5.0%; Score 202.5; DB 16; Length 2174;
Best Local Similarity	22.7%; Pred. No. 0.0052;
Matches 176; Conservative 90; Mismatches 312; Indels 197; Gaps	

QY	12	AIGGGFVDCIVFNEGAPCILLVRT-DTGGMYRWDAANGRWIPLLDNVGNHMCYNCGVS	70
DB	976	SIGGG-----GNGGVGSNKEIGSGFNLTA-----NVGVGGSGS	1011
QY	71	IADPINTNKVAAAVGYMTNSDPNDGAILRSSDQA-----TWQTILPFK	117
DB	1012	GGASGN----AIVGL-----DSGTHLQTSKGARGVIQSIGGGGTSQGASVLS	1059
QY	118	LGNMFGRMGERLAVDPPNNILIFGPAGSKGLWRSTDSCATWSQNTNFPDVGTYTANP	177
DB	1060	ASASLPGGG--EEAAEAEESEGSGAFSAVGSVGRGTGGSGS-SGTVNVTACT----	1112
QY	178	TDTTGYSDDIQGVVMVAFDKS---SSSLGAOSKTIFYCVADPNPVWSRDG-----	226
DB	1113	--ISTFGADADQLAQSIGGGGLGSGVQASGDSEPDDEGSECAGEGNGDGHGY	1170
QY	227	-----GATWAQVAPCAPTGFIPHKGVDFPNHVLIATS-----NTGGPYDGSSG	270
DB	1171	CFGVSVGAT--IDDGCTGTAAANGNAVTLTHAGIATAGWDAGIVAQISGG--GGAG	1225
QY	271	DWKFSVTSWTTRISPVPSTOTANDYFGYSLGITDRQHPNTINVAQTISWPDTIFRS	330
DB	1226	GTSTAGSQATANIITVGGSGCAGGNGAVGITFDDNHGNSISTAGYSAY--GVLLOS	1282
QY	331	TDSGATWTRIWDWTSYPNBSLAYLD-----ISAEPWLTFGVQPNPVPVS	375
DB	1283	IGGGG-----QQGDGSDAAGRITVGGGFGSGGAGSGGMVTKGININSTG-----	1332
QY	376	PKLGWMDEAMAI--DPNFSDRMLYGTGATLYA-----TNDLTKWDSGGQIHAPMVKGL	427
DB	1333	-----DDAHGIVAQSIGGGGVGGAGSGSTAEREHSHTIDLTVGGSGG----VGGSGGE	1382
QY	428	EETAVNDLISPSPGALISAL-----GDLGCFHADVTAVPTIFTSPVTTGTSVDY	480
DB	1383	VDSLVTGTTLS-TSGARALGLVAQSIGGGGGIGGAGEAD--SIASLVVGG--SGGGTIDG	1436
QY	481	AELNPSIIVRACSFDPSPQNRDH--VAFSTDGKN-----WFQSEPQGVVTGG	528
DB	1437	GAVTYDLTQSOST--TTQGIHAHLVAQSTGGGGVGGGAASGAPLFTGNSPGSYDGG	1493
QY	529	TVAASADGREFV--WAPCDGPQVVYAVGFENSWAASQGVPAANAIRSDRVNPKTFYAL	585
DB	1494	DVAVTAGSLFRGDYAFGLAQSTGGGGGREGN-ATSAFTGSNGNLUSDG-----	1543
QY	586	SNGETFYRSTDGGVTFOPVA-----AGLPSSGAVGYMFH-AVPKREGD----LWL	629
DB	1544	KSGNVTVSLDAGRTIOASCDSIGIFAQSDAGTDNNGTIDTVNCTVGTGSGDNGAGIW	1603
QY	630	AASGLYLHSTNGGSSWALTGVSSAVNWVFCKSPGSSYPAVFVGTI-GGVTA	683
DB	1604	SAGKDNIWTVNSGGNYSAASGV--AVQYTAGNNSPEDTLVNNAGTISGSVKGA	1656

RESULT 10
P73139 PRELIMINARY:
ID P73139 PRS: 3972 AA;

AC	P73139;	
AD	01-FEB-1997 (TrEMBLrel. 02, Created)	
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)	
DE	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	HYPOTHEICAL 418.3 KDA PROTEIN.	
GN	SLR1028	
OS	Synechocystis sp. (strain PCC 6803).	
OS	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.	
OX	NCBI_TaxId=1148;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=97061201; PubMed=8905231;	
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,	
RA	Miyajima N., Hirosewa M., Sugliura M., Sasamoto S., Kimura T.,	
RA	Hisouchi T., Hirasawa A., Muraki A., Nakazaki N., Naruo K., Okumura S.,	
RA	Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,	
RA	Tabata S.;	
RT	"Sequence analysis of the genome of the unicellular cyanobacterium	
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the	
RT	entire genome and assignment of potential protein-coding regions.;"	
RL	DNA Res. 3:109-136(1996).	
DR	EMBL; D90904; BAA17165.1; -.	
DR	HSP; P30617; LJPC.	
DR	InterPro; IPR001480; B_lectin.	
DR	InterPro; IPR003644; Calx_beta.	
DR	InterPro; IPR000413; Integrin_alpha.	
DR	InterPro; IPR000215; Serpin.	
DR	Pfam; PF01453; Agglutinin; 1.	
DR	Pfam; PF01839; FG-GAP; 8.	
DR	PRINTS; PR01185; INTEGRINA.	
DR	SMART; SM00108; B_lectin; 1.	
DR	SMART; SM00237; Calx_beta; 1.	
DR	SMART; SM00191; Int_alpha; 9.	
DR	PROSITE; PS00284; SERPIN; UNKNOWN_1.	
KW	Hypothetical protein; Complete proteome.	
SO	SEQUENCE 3972 AA: 418281 MW: B9F849F274FBEABD CRC64;	

Query Match 4.7%; Score 189; DB 16; Length 3972;

Best Local Similarity	19.3%;	Pred. No.	0.063;
Matches	166;	Conservative	93;
Mismatches	166;	Mismatches	276;
Indels	326;	Indels	326;
Gaps	42;	Gaps	42;

Qy	13	IGGGGFVDGIVFNEGAPGILYVTRDIGGMWRDAANG-----RWIPLLDW--	57
Dd	2458	VNGDGFADVISGGSPAGGVLIIFGNSTKDLL---DAALGTDDLIIISVENAQVKFEVALGDFFD	2515
Qy	58	-----VGWNWCYNG-----VVSIAADPINTNKVAAVGMVNSWD--	93
Dd	2516	GDGLADGFVIDDOGNFELVLGSPELGSQSLVIDSTLPNLSNFNOAWGVDFENGVDFF	2575
Qy	94	----PNDDGAILRSSDOGATWQTITPFKLGGNNP-----GRGMGERLAVDPN--	136
Dd	2576	VLOGPNSTIAVYGNANGTLTDSPLTF--GNNFPSPSFTGIDLNGCIKEIVAGQPPLN	2633
Qy	137	-----NDNIIFYCAPSGCKGLWRST-----DSG-ATWSQMTNFDPDVGYTANP	177
Dd	2634	PVPNIGFGGGLQFYTYEAGNAVLOPTYNPNNASVTEASGLSSWGQI-SFPNQYAQAQVP	2692
Qy	178	TDTT--GY-----QSDIQGVVV-----VAFDKSS-----	200
Dd	2693	SFATLDGLWLQAEYGINERISTKDSYIYIORSRDGVSWENLTQVVPDLSNGTPIDLKNL	2752
Qy	201	-SLQASKTTIFVGVADPNPVFWSRDG-----GATQOAVP-----GAPT--GFPHK	244
Dd	2753	PSITAYNGTLYLGFTADNGOV-WVAEGVNTNANSGLINAVPIQNASNNGPTLVAFNDEL	2811
Qy	245	GVE----DPNVHVIYLIATSNTGPYDGGSDGVKFSVTGWTRISPSTDTANDFGY	300
Dd	2812	YVFVKDASDNLDILYSNSSNPG-----SSSG--WDGSTVLFESDVNQATN-----FPL	2858
Qy	301	SGLTIDRQHPTINMVAQTISWMPPDTTIIFRSTDGGATWTRI-----WD-----	342
Dd	2859	SATVPCPLCDGTTLAVA-----FRSNNSPATWYGILLNSSDVTNWOGSAELTOVDA	2907

RESULT 10

Р73139

ID P73139

ID P73139 PRELIMINARY;

PRT; 3972 AA.


```
Q9N1P0
ID Q9N1P0 PRELIMINARY; PRT; 818 AA.
AC Q9N1P0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE SUBMAXILLARY MUCIN (FRAGMENT).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BREED ANGUS; PubMed=10759843;
RX MEDLINE=20223253;
RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts.";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL; AF178428; AAF67279.1; -.
FT NON_TER 1
FT NON_TER 818
FT SEQUENCE 818 AA; 73014 MW; B5B44F84F66F86B CRC64;
SQ

Query Match 4.6%; Score 185.5; DB 6; Length 818;
Best Local Similarity 25.9%; Pred. No. 0.014;
Matches 159; Conservative 55; Mismatches 249; Indels 151; Gaps 34;

QY 120 GNMPPGRMGERLAVDNNNNILYFGA-----PSGKLMRSTDSGTATWSQMTNFPDVGTY 173
DB 182 GSSPGRSRATAVSGESQPTVALSGATGTSAGPSGT---RSTSSAIP----- 225

QY 174 IANPTDTTGQSDIQGVVWVAPD-KSSSLGQASK-TIFVGVADPNPNFVSRD---GG 227
DB 226 -ATPGTTGRAA---GAGTAVDSQQTASLPAAARTALPGTSPAGTSESRSVPGG 281

QY 228 ATWQAVPGA-----PT-----GFIPKGVDFDPVNHVL--YTATNTG---G 263
DB 282 SETTQPGAGSEPTLSPGVTRTALRGSETPVSTGVSLPGSTOGGSAATGSGAGSG 341

QY 264 PYDSSGDVWKFVSGTWTRISPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWNP 323
DB 342 PTAPVSGET-RTSVISGTVNVPVSCAPVTPGSS--AGSSGA-----PGT-----GGP 384

QY 324 DTIIFRSTOGGATWTRIDWTSYPNRSLRYVLDISAE-PWLTFGVQPNPVPSPKLGWMD 382
DB 385 GSETASPLSAAATGATGSGTSLPPSG---APVTPEPPLISTGASAGPPASSESTVTL 440

QY 383 EAMAIIDPFNSDRMLYGTGAPLYATNDLTWKDSGGQIHAPMVKGLEETAVNDLISPPSG- 441
DB 441 GATGTDVLR-----GTSLPV-----SGGAVTPAPSPGSSATA-----GPCVGS 480

QY 442 APLISALGDLGGTHADVTAAP-STIFTSVFTTGTSDVYAEINPSIIVRAGSDFSSQP 500
DB 481 ATTQVQSGATGADVLRSGTSLPVSGVAVSPGSSPG-----RSGTAVSS-- 524

QY 501 NDRHVFSTDGKNWFGSPGVTGGT-VAASADGSRFVWA--PCDPCQPVVYAVGFG 557
DB 525 -----OGSQTVALSGATGTSVPGSTREFSSAIPATPGSTTGRAAG 568

QY 558 NSWAAASO---GVPANAIQRDRVNPRTFYALSNGTFFRST-DGG--VTFQPVAAAGLPSG 611
DB 569 TPGVDSQQTASLPAA--RTALPGTSPAGTSESRSVPVGGSETTQPGAGSEPTL 626

QY 612 AVGV-MFHAVPGEGDLWLAAASHSTNGGSSWSAITGVSSAVNVGFGKSPGSSYPA 670
DB 627 SPGVTRTALRGSETPVSTGVSLPGSTOGG---SAATGSGA---GSGPTAPVSGETR 680

QY 671 VFVV-GTIGGVTGA 683
```

```
Db 681 TSVISGTVNVPVSGA 694

RESULT 13
Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CELLULASE PRECURSOR (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00598; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1
FT SIGNAL 34
FT SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;
SQ

Query Match 4.6%; Score 185.5; DB 2; Length 997;
Best Local Similarity 21.1%; Pred. No. 0.018;
Matches 165; Conservative 85; Mismatches 248; Indels 285; Gaps 43;

QY 79 NKVMAVAGMTNS---WDPNDGAILRSSDQOGATWQITPLPFKLGNNPGRMGRLAVDP 135
DB 147 NELWQGVGAGTNDHAWGPAE-----VQMNRPSFKIDASCPSGDLAAETAAL 195

QY 136 NDNILYFGAPSGKGLWRSTDSGTWS-----OMTNEPDV--GTYIANPTDT----- 180
DB 196 AASIVF-----ADSDPVYSAKLLQHAKELYNFADTYRGKYTCIIDAFAFYN 243

QY 181 --TGYQSDIQ-GVWV-----AFDKSSSLGQASKTIFVGVADPNPNPVFV 222
DB 244 SWTGYEDELAWGAWLYLATNDNAYLSKALSAADRWSTSGSA-----NWPYTW 292

QY 223 SRDGGATWQAVPGAPTGFIPHKGVDFDPVNHVLIATNTGPGDYDSSGDVWKFVSVTSGTW 282
DB 293 TQG-----WDS-----RHYG-----AQILLARITSLNMP-----EATKF----- 322

QY 283 TRSPVPSTDTANDYF--GYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGGATWTRI 340
DB 323 -----IQSTERNLDYTVTGTNGGRV-----KTPPGGLAWLDQ 354

QY 341 W-----DWTSTPNRSLRYVLDISAEFPWLTFGVQPNPVPSPKLG----- 379
DB 355 WGSRLRYAANAATFSFVYSDNWVSDPVKRSRY--QNFAISQINYILGDNPROSSVYVVGQGN 412
```



```
QY 380 -----WMDAMAIDPNSDRMLYGT-----GATLYATNDL-----TKW 412
Db 413 SPQPHRTAHSWMNEDI--PANHRHILYGAMVGPNASDQYTDIGDYVSNVATDY 470
QY 413 DSG-----GOIH-----TAPMVKGLE-----ETAVNDLISPPSG 441
Db 471 NAGFTGALAKMNLGYGNHQPANFPAPEVKGYEYFVEAAVRSSGNSYTEIRALLNRS 530
QY 442 APLISALDGLGGFTH---ADVTAVSTFTSPTFTGTSVDYAEINPSIIVRAG----- 492
Db 531 WP--ARMGDQLSKYFLDLSEVTAAGRTV--SDVQVTVSSSEGAISQPVVVDAAKRIYA 586
QY 493 -----SPDSSQPNDR-HVAFSTDGKKN-WFQSGEPG--GVTTGG----- 528
Db 587 ITANESNTKIYPGEGNYRKEVQFRITGPOGANPANDPSYQNLTTGNPKVSNYIPYDA 646
QY 529 --TVAASADGRFRVWAPDGPQPVVAVGFGNS-----WAAASGVFANAQIRSD-RVNP 580
Db 647 GKVSGQEPGLVTPVAVPAAPAG--VOAVA--GNSQVALNWSASAGAVSYTVKRAEVS 703
QY 581 TFVALS-NGFYRST--DGGVTFQPVAAGL-----PSSGAVGVMFHAVPGKEGDLW 631
Db 704 TTVAAVNGLTYNTGLTNGKTYIYVTVAVNSAGESPASVQVSGMPQAAATTVPCAV 763
QY 632 SSGLYHSTNGGSSWASITGVSSAVNVGFGKSPGSSYPVAVFV-----VGTIGGVTG 682
Db 764 TAG---NNQNLWSATAAGSASTY--VQRAVAGTYTVDATGLAVLNYNDTTALNGTSY 817
QY 683 AYR 685
Db 818 SYR 820

RESULT 14
Q9HLQ9 PRELIMINARY; PRT; 998 AA.
AC Q9HLQ9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE THERMOPHILIN PRECURSOR RELATED PROTEIN.
GN TA0167.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445063; CAC11313.1; -.
DR MEROPS; A05.0PW; -.
KW Complete proteome.
SQ SEQUENCE 998 AA; 108669 MW; EC2E9E92E382853 CRC64;
```

```
Query Match 4.6%; Score 184.5; DB 17; Length 998;
Best Local Similarity 20.1%; Pred. No. 0.02;
Matches 173; Conservative 113; Mismatches 285; Indels 289; Gaps 47;

QY 9 SNVAIGGGGFGVDGIVFNEGAPGI-----LYVRTD-IGG----MYRWDAAANGRWIP- 53
Db 188 ANALHGNGKLVPGVYFYDGPTEFNISFPFSLKLYLNSSLVGGNDAYFNYSIENGKITKS 247
QY 54 -LLDWYCWNNWYGVVSTAADPINTNKVWAAVGMVTNSHDPNDGAILRSSDQATWQIT 112
Db 248 GSYDAVTFNSSG-----KASSPARFLISGYSPSPAGLLYDAELAITG--- 289
```

```
QY 113 PLPFLKGGNMPGRGMRGERLAVDPNNDNLYFGAPSGKGLWRST-----DSGATW 161
Db 290 -----PGGG-----SNVNYGICNGTGLYLYNSTSKAFNTVRSAYDAGVDT 330
QY 162 SQTNPFDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFGVGADPNPVF 221
Db 331 GETSYGVDVGWY--GNMADLQSGPSLIYG-LW-----NVSSGIETI-SGTAEPQSYIF 379
QY 222 WS-----ROGGATWQAVPGAPTGFPHKGVDPVNHVL-----YIATSNTPGPDGSS 269
Db 380 VSNKGFNNSTASWAPYSGAHFRKLPKGSYSLA--VLRNLYAPEYINDAGTSPSLNGTS 437
QY 270 --GDVWK-----FSVTSGTWTTRISPV-----PSTDANDYF--GYSG 303
Db 438 KTGDIVAPVIENYSDALYYSRSGNSQISPFVIFSGQTVSVDPFLGVLNDLYFPVFTGL 497
QY 304 T-----IDRQHPNTIMVATQISWNPDTIIFRSTGGGATWT-RIMDWTSYPNRSRYVL 355
Db 498 AIINTSYHIDVENASSEFEITYGSQY--DSLASHGLGNTNMGMILYGTG--NVSITGGL 553
QY 356 DISAEPWLTFGVQPNPPVSPKLGWDEAM-----AIDPFNSDRMLYGTGATL 403
Db 554 NIT-----GWPEPMKGFVAVANLWVWNSNDLIADNVFI-----SL 589
QY 404 YATNDLTKWDSGGQIHIAPMVKGLEETAVND-----LISPPSGAPL-----ISALGD 450
Db 590 HASSS---DLOPPYSSVLLIYGMNSTRANTVWGNVFPDQYVSENGNATVGLFVDSSGN 645
QY 451 L---GGFTHADVTAVPS-TIFTSPTVTTGTSVDYAEINP---SIIVRAGSDFDSSQPNDR 503
Db 646 LIYNNAFTQYDIAFSPDFNIYTDNATYHDNWNISSPEPINYINVGYSL-----NGS 699
QY 504 HVAFTSDGKKNWFGSGEPGGVTTGGTVAASADGSRFWAP-----GDPQPVV--- 551
Db 700 IVGGYEGGNYW-----GSPITPWNESGYIDSYDHPLYYFL 739
QY 552 YAVFGNSWAAASQGVPPANA--QIRSDRVNPKTFYALSNGTFYRSTD---GGVTFQPVAA 606
Db 740 YELSF-----AIHGLPAGTSWELVADGV-----YVVGSGPYINITEFNFGNTTY---AVF 786
QY 607 LP-----SSG-----AVGVMEHAVPGKEGDLW-LAASSGLYHSTN 640
Db 787 LPPGYISNAGSAYAYATDMSTVNYARHVMYTGIVFHTATNYVKNETWYLDIAGLSYSTS 846
QY 641 GGSWSAITGVSSAVNVGFKSGPSYPAVFVVGITGGVT-----GAYR----- 685
Db 847 QDLTVSLPNGTYS-----YSVFAPPNNYTAVSGYVTVNGVSEIVNITFPQLRYVIRET 901
QY 686 SDDCGTTWVLINDDQHQYGN 705
Db 902 GLPAGSSWYVVIDGKAYYSN 921

RESULT 15
Q9RK65 PRELIMINARY; PRT; 848 AA.
ID Q9RK65;
AC Q9RK65;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SCF11.25.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:50 ; Search time 101.03 Seconds
(without alignments)
97.848 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231.5	49.3	782	AA15625	Cellulase AE-1. A
2	195.5	41.6	616	AA13494	Truncated cellulase
3	194.5	41.4	1751	AA13493	Truncated cellulase
4	191.5	40.7	1426	AA13492	Truncated cellulase
5	186	39.6	499	AA42122	NK-1 cellulase. B
6	172	36.6	700	AA13227	Novel endoglucanase
7	165.5	35.2	551	AA18790	Corrected Bacillus
8	165	35.1	167	AA195080	Cellulose binding
9	165	35.1	476	AA154123	A mannanase-linker
10	165	35.1	493	AA228850	Pectate lyase-linker
11	165	35.1	493	AA43218	Pectate lyase CBD

12	165	35.1	531	18	AAW15238	Scaffoldin protein
13	165	35.1	1853	19	AAW43108	C. thermocellum ce
14	155.5	33.1	1352	22	AAG63962	Amino acid sequenc
15	150.5	32.0	1350	22	AAG63963	Amino acid sequenc
16	120.5	25.6	531	16	AAW01503	60 kD endoglucanas
17	120.5	25.6	532	12	AA13229	Endoglucanase enco
18	114	24.3	162	15	AA163634	Cellulose binding
19	114	24.3	162	20	AAW90077	C. cellulovorans C
20	114	24.3	163	22	AAE05745	Clostridium cellu
21	114	24.3	256	22	AAE05745	Clostridium cellu
22	114	24.3	328	22	AAE05749	Chimeric S peptide
23	114	24.3	341	22	AAE05747	Clostridium cellu
24	114	24.3	428	22	AAE05748	Clostridium cellu
25	113	24.0	190	22	AAE05746	Clostridium cellu
26	113	24.0	382	20	AA139952	Gaussia luciferase
27	112	23.8	154	20	AAW90081	C. cellulovorans C
28	112	23.8	156	20	AAW90080	C. cellulovorans C
29	75.5	16.1	986	21	AA140440	Human brain-derive
30	75.5	16.1	1346	22	AAU04567	Human G-protein co
31	75.5	16.1	1346	22	AAU04581	Human G-protein co
32	75.5	16.1	1371	22	AAU04570	Novel human diagno
33	75	16.0	1723	22	ABG24680	Novel human respir
34	72.5	15.4	69	22	AAU17759	Fragment F1029 of
35	70.5	15.0	2965	19	AAW56450	Fragment HJ1916 o
36	69	14.7	2970	19	AAW56445	Fragment HJ1916 o
37	68	14.5	2873	19	AAW56441	Nasturtium xyloglu
38	67.5	14.4	857	17	AA195287	Sequence of the si
39	67	14.3	236	15	AA145442	Anti-erbB2 scfv.
40	66	14.0	237	17	AA194020	Single-chain anti-
41	66	14.0	237	18	AAW15185	Human p53 regulato
42	65.5	13.9	1566	20	AA106309	Protein sequence o
43	64	13.6	117	20	AAW86130	Murine monoclonal
44	64	13.6	143	20	AA150155	Drosophila melanog
45	64	13.6	254	22	ABB64503	

ALIGNMENTS

RESULT 1

AA15625	AA15625 standard; Protein; 782 AA.
ID	AA15625 standard; Protein; 782 AA.
XX	AA15625;
AC	AA15625;
XX	AA15625;
DT	17-MAR-1992 (first entry)
XX	Cellulase AE-1.
DE	Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
KW	Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
XX	Aeromonas strain no. 212.
OS	JP03251174-A.
PN	08-NOV-1991.
PD	08-NOV-1991.
XX	28-FEB-1990; 90JP-0045465.
PF	28-FEB-1990; 90JP-0045465.
XX	(OJIP) OJI PAPER KK.
XX	WPI; 1991-373412/51.
DR	N-PSDB; AAQ15178.
XX	Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
PT	of opt. pH when carboxymethylcellulose is used as substrate.
XX	Claim 2; Fig 3; 8pp; Japanese.
PS	The sequence was deduced from the gene which was sequenced from
XX	plasmid, pAEC 1, prepd. by ligating chromosomal DNA contg. the
CC	

gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt. of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-methylcellulose is the substrate. The N-terminal sequence: GIHADT- has been confirmed by Edman degradation. The gene can be used to produce recombinant enzyme which is used for the effective utilisation of biomass resources and the mfr. of pharmaceuticals and foodstuffs, and also for the detergent and deinking of waste paper.

XX

SQ Sequence 782 AA;

Query Match 49.3%; Score 231.5; DB 12; Length 782;
Best Local Similarity 48.3%; Pred. No. 5e-18;
Matches 42; Conservative 16; Mismatches 28; Indels 1; Gaps 1;

QY 2 SGGVKVOYKKNDSAPGDNQIKPGQLVNTGSSVDLSVTVRVWFTRDGGSTLVYNCDW 61
||| ||| | : ||| || : ||| || : ||| || : ||| || : ||| ||
Db 631 sgdlaydykgdgtnaadnqfphfnvknkaavplseisrlyftad-gndqiyncdw 689

QY 62 AAMCGNIRASFSGSVNPATPTADTYLQ 88
| : || | : : : || ||| ||:
Db 690 amvcslngafvkmpgkanactyle 716

RESULT 2
AAV13494
ID AAV13494 standard; Protein; 616 AA.
XX
AC AAV13494;
XX
XX 30-JUL-1999 (first entry)
XX
XX Truncated cellulase Cel E3/B5.
DE
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel I/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
XX Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
XX 15-SEP-1998; 98EP-0810919.
PF
XX 19-SEP-1997; 97US-0932571.
PR
XX (CLRN) CLARIANT FINANCE BVI LTD.
XX PA
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI; 1999-315403/27.
DR N-PSDB; AAX55660.
XX
XX New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
PT
XX
XX Claim 7; Page 42-43; 65pp; English.
XX

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, and Cel E1, Cel E1/2, Cel I/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new enzymes are useful in laundry detergent compositions to prevent or

KW	Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW	endo-1,4-mannanase; Bacillus sp. I633; galactomannan;
KW	1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW	galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW	printing paste; plant material degradation; recycled waste paper;
KW	paper making pulp; guar; locust bean gum; thickener; viscosity;
KW	mannan-containing food; coffee extract; cleaning composition;
KW	machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW	contact lens; body-care composition; fabric softener; oil well drilling
KW	subterranean formation fracture; cellulose binding domain.
XX	
OS	Synthetic.
OS	Bacillus sp.
OS	Clostridium thermocellum.
XX	
XX	WO9964619-A2.
PN	
XX	
PD	16-DEC-1999.
XX	
PF	10-JUN-1999; 99WO-DK00314.
XX	
PR	10-JUN-1998; 98US-0111256.
PR	20-OCT-1998; 98DK-0001340.
PR	20-OCT-1998; 98DK-0001341.
PR	28-OCT-1998; 98US-0105970.
PR	28-OCT-1998; 98US-0106054.
PR	23-DEC-1998; 98DK-0001725.
PR	05-MAR-1999; 99DK-0000306.
PR	05-MAR-1999; 99DK-0000307.
PR	05-MAR-1999; 99DK-0000308.
PR	05-MAR-1999; 99DK-0000309.
PR	09-MAR-1999; 99US-0123543.
PR	10-MAR-1999; 99US-0123623.
PR	10-MAR-1999; 99US-0123641.
PR	11-MAR-1999; 99US-0123642.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX	
DR	WPI; 2000-105891/09.
DR	N-PSDB; AAZ45336.
XX	
PT	New mannanases for treatment of textiles, plant material and coffee
PT	extract, and in cleaning compositions -
XX	
PS	Example 4; Page 211-212; 242pp; English.
XX	
CC	The present sequence represents a mannanase-linker-cellulose binding
CC	domain fusion protein. Mannanase (also known as mannan
CC	endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC	hydrolyses galactomannans. Specifically, mannanases hydrolyse
CC	1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans
CC	and galactoglucomannans. The mannanase protein, or preparations
CC	containing it, are used to improve properties of cellulosic or
CC	synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC	sizes or printing pastes). They are also used to degrade or modify
CC	plant materials (particularly recycled waste paper, paper making pulps,
CC	or material containing guar or locust bean gums (thickeners), or to
CC	reduce viscosity of mannan-containing foods or feeds). The mannanases
CC	are also used to process coffee extracts (to inhibit gel formation);
CC	in cleaning compositions (for machine washing of fabrics, as
CC	hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC	dental, contact lens or body-care compositions) where they remove
CC	mannan-containing soils and prevent binding of some soils to
CC	cellulotics; and in fabric softeners. They can also be used in oil
CC	well drilling to fracture subterranean formations.
XX	
SQ	Sequence 476 AA;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSSSTLVYVNC 60

Db 314 vsgnlkvfynpsdttnsinpqfvtngssaidskltlryyytvdgkqdtfw-cd 372

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 373 haaligsngyngitsnvkgtfvkmsstnnadtyle 409

RESULT 10

AAAY28850
ID AAY28850 standard; Protein; 493 AA.

XX AC AAY28850;

XX DT 17-JAN-2000 (first entry)

XX DE Pectate lyase-linker-CBD fusion protein.

XX KW Pectate lyase-linker-CBD fusion protein; ATCC 14580; plasmid pMB914;
KW cellulose binding domain; CBD; linker; transformed; detergent;
KW cellulosic fibre; yarn; degradation; recycled waste paper; animal feed;
KW paper-making pulp; retting process; processing; wine; juice.

XX OS Chimeric - Bacillus licheniformis.

XX OS Chimeric - Clostridium thermocellum.

XX PN WO927083-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00514.

XX PR 24-NOV-1997; 97DK-0001344.

XX PR 06-MAY-1998; 98US-0073684.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;

XX DR WPI; 1999-610578/52.

XX DR N-PSDB; AAX90978.

XX PT New isolated pectate lyase enzymes -

XX PS Example 4; Page 85-86; 93pp; English.

XX CC The present sequence is a Pectate lyase-linker-CBD fusion protein.

XX CC Plasmid pMB914 was constructed using pectate lyase gene from

XX CC Bacillus licheniformis, ATCC 14580 and cellulose binding domain (CBD)

XX CC gene from Clostridium thermocellum YS fused with a linker molecule.

XX CC Bacillus subtilis was transformed with plasmid pMB914 for expression of

XX CC the fusion protein. Pectate lyase can be used in detergent compositions,

XX CC for cleaning hard surfaces, for machine treatment of fabrics, for

XX CC improving the properties of cellulosic fibres, yarn, woven or non-woven

XX CC fabric, for the degradation of plant material e.g. recycled waste paper,

XX CC mechanical paper-making pulps or fibres subjected to retting process, for

XX CC preparing animal feed and for processing wine or juice.

XX SQ Sequence 493 AA;

Query Match 35.1%; Score 165; DB 20; Length 493;

Best Local Similarity 36.1%; Pred. No. 1.4e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSSSTLVYVNC 60

Db 331 vsgnlkvfynpsdttnsinpqfvtngssaidskltlryyytvdgkqdtfw-cd 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 390 haaligsngyngitsnvkgtfvkmsstnnadtyle 426

RESULT 11

AAAY43218
ID AAY43218 standard; Protein; 493 AA.

XX AC AAY43218;

XX DT 13-JAN-2000 (first entry)

XX DE Pectate lyase CBD fusion protein sequence.

XX KW Pectate lyase; polysaccharide lyase; enzyme; pectin degradation;
KW polygalacturonide; detergent composition; hard surface treatment;
KW cellulosic fibre; plant material degradation; recycled waste paper;
KW mechanical paper-making pulp; wine processing; cellulose binding domain;
KW CipB.

XX OS Clostridium thermocellum.

XX OS Bacillus sp.

XX OS Synthetic.

XX PN WO927084-A1.

XX PD 03-JUN-1999.

XX PF 24-NOV-1998; 98WO-DK00515.

XX PR 24-NOV-1997; 97DK-0001343.

XX PR 24-NOV-1997; 97DK-0001344.

XX PR 06-MAY-1998; 98US-0073684.

XX PR 02-NOV-1998; 98US-0184217.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Andersen LN, Schuelein M, Lange NEK, Bjornvad ME, Moller S;

XX PI Glad SOS, Kauppinen MS, Schnorr K, Kongsbaek L;

XX DR WPI; 1999-610579/52.

XX DR N-PSDB; AAX31562.

XX PT New isolated pectate lyase enzymes -

XX PS Claim 26; Page 106-108; 113pp; English.

XX CC This sequence is a fusion protein comprising a Bacillus species
XX CC pectate lyase of the invention, fused via a linker to the Clostridium
XX CC thermocellum CipB protein cellulose binding domain (CBD). The pectate
XX CC lyase enzymes are obtained from Bacillus agaradhaerens (BA), Bacillus
XX CC licheniformis (BL), Bacillus halodurans (BH) and other Bacillus species.
XX CC The pectate lyase enzymes can be used for degrading pectin, pectate and
XX CC polygalacturonides. They can be used in detergent compositions, for
XX CC cleaning hard surfaces, for machine treatment of fabrics, for improving
XX CC the properties of cellulosic fibres, yarn, woven or non-woven fabric, for
XX CC the degradation of plant material (e.g. recycled waste paper, mechanical
XX CC paper-making pulps or fibres subjected to a retting process, for
XX CC preparing animal feed or for processing wine or juice. DNA encoding the
XX CC enzymes can also be used for the production of transgenic plants.

XX SQ Sequence 493 AA;

Query Match 35.1%; Score 165; DB 20; Length 493;

Best Local Similarity 36.1%; Pred. No. 1.4e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNOIKPGLQLVNTGSSSVDLSTVTVRWFTRDGSSSTLVYVNC 60

Db 331 vsgnlkvfynpsdttnsinpqfvtngssaidskltlryyytvdgkqdtfw-cd 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 390 haaigsgnygitsnkvgtfvmssstnnadtyle 426
 II:: :I :: :I |||||:
 Query Match 35.1%; Score 165; DB 18; Length 531;
 Best Local Similarity 36.1%; Pred. No. 1.5e-10;
 Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps

AAW15238 standard; protein; 531 AA.
 AC AAW15238;
 XX 1 VSGGVKVOYKNNDSPAGDNQIKPGLQLVNTGSSVDLSVTYVRYFTRDGGSSSTLYVNC D 60
 III::III:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I:: I::
 Db 312 vsgnlkvefygnpsdttinsnpqfkvtntgssaidlsklrtlyryyvdgkdqdtfw-cd 370

XX 28-JAN-1998 (first entry)
 DE Scaffoldin protein from Clostridium thermocellum.

Dockerin; Celd; CIPa; scaffoldin; cellulose binding domain;
 chromatographic separation; soluble substrate modification; CBD;
 multi-enzyme delivery system; animal feed; paper production;
 plant protection; pest control.

Clostridium thermocellum.
 OS
 XX Key Location/Qualifiers
 FH Region 1..153
 FT /label= internal_repeat_element_1
 FT Region 134..306
 FT /label= internal_repeat_element_2
 FT Domain 239..531
 FT /label= cellulose-binding
 FT /note= "only 60% of the CBD is present"
 XX WO9714789-A2.

PN 24-APR-1997.
 PD 16-OCT-1996; 96WO-US16485.
 PF 17-NOV-1995; 95US-0559968.
 PR 17-OCT-1995; 95US-0005701.
 XX (GENEV) GENENCOR INT INC.
 PA Bott RR, Clarkson KA, Fowler T, Liu C, Ward M, Xia H;
 PI WPI: 1997-245106/22.
 DR Composition with enzymes non-covalently bound to a peptide backbone
 PT - used as a multi-enzyme delivery system, e.g. in food processing,
 PT textiles and pest control
 XX Claim 10; Fig 6; 29pp; English.

This protein sequence represents a scaffoldin based on the CipA protein
 of Clostridium thermocellum. The scaffoldin protein is used in a novel
 composition that comprises at least 2 enzymes non-covalently bound to a
 peptide backbone (i.e. present sequence). The scaffoldin comprises a
 number of internal repeating units and at least one cellulose binding
 domain (CBD). The CBD may be altered to modify its affinity for
 cellulose, which may be desirable where cellulose binding would be
 disadvantageous. The enzyme is bound to the scaffoldin by a dockerin
 region of the enzyme, which binds to a repeating element of the
 scaffoldin. The dockerin is preferably Cells (AAW15237) or Celd
 (AAW15236). The composition can be used in reducing allergenicity,
 producing synergistic effects, and facilitating selective modification
 of substrate. By taking advantage of the cellulose binding domain of the
 complex, the complex could be immobilised for use in chromatographic
 separations or for soluble substrate modification. By adding the
 scaffolding domain, it is possible to recover enzymes, or to quantify
 the amount of an enzyme in a solution. The composition could also be
 used in a multi-enzyme delivery system which could be used in the food
 industry, in food processing, animal feed, textiles, bioconversion,
 pulp and paper production, plant protection and pest control, as a wood
 preservative, topical lotions, and biomass conversions.

QY 4 GVKVQYKNNDSAPGDNQIRPGLQLVNTGSSVDLSTVTYVWFTRDGSGSSTLVYNCDMAA 63
Db 1199 givlgvrtadtnvndhlnphqilnkgfistvpinelkiryvyytidgdreg-tfncdyav 1257
QY 64 MCGGNIRASFGSVNPNATPTADTVLQ 88
Db 1258 iscskingklvkmkkaatgadyyle 1282

Search completed: July 2, 2002, 09:15:51
Job time: 239 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:54 ; Search time 39.87 Seconds
(without alignments)
54.524 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGKVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	616	4	US-09-136-574A-47
2	194.5	41.4	1751	4	US-09-136-574A-44
3	191.5	40.7	1426	4	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PCT-US95-13813-9
7	165	35.1	493	4	US-09-198-956-10
8	165	35.1	493	4	US-09-198-955A-12
9	120.5	25.6	531	2	US-07-862-588B-7
10	114	24.3	162	1	US-08-048-164A-2
11	114	24.3	162	1	US-08-460-462-2
12	114	24.3	162	1	US-08-460-457-2
13	114	24.3	162	1	US-08-460-458-2
14	114	24.3	162	2	US-08-460-455-2
15	114	24.3	162	2	US-08-330-394A-2
16	114	24.3	163	3	US-09-006-636-7
17	114	24.3	163	4	US-09-006-632-7
18	113	24.0	382	4	US-09-277-716-22
19	112	23.8	154	2	US-08-330-394A-29
20	112	23.8	156	2	US-08-330-394A-22
21	64	13.6	428	3	US-09-118-319-5
22	64	13.6	464	1	US-08-353-400-36
23	63.5	13.5	1785	4	US-09-341-587-3
24	63	13.4	1581	4	US-09-110-517-2
25	62.5	13.3	288	4	US-09-423-439-38
26	62.5	13.3	445	1	US-08-353-400-33
27	62.5	13.3	673	4	US-09-423-439-32

28	62.5	13.3	802	4	US-09-081-345-18	Sequence 18, Appl
29	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appl
30	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
31	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
32	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
33	61.5	13.1	307	4	US-08-370-476-68	Sequence 68, Appl
34	61.5	13.1	316	1	US-08-260-202A-18	Sequence 18, Appl
35	61.5	13.1	316	1	US-08-017-114-18	Sequence 18, Appl
36	61.5	13.1	316	3	US-08-505-307-18	Sequence 18, Appl
37	61.5	13.1	316	5	PCT-US94-02034-18	Sequence 18, Appl
38	61	13.0	301	2	US-08-656-906-25	Sequence 25, Appl
39	61	13.0	301	4	US-09-217-847-25	Sequence 25, Appl
40	61	13.0	617	1	US-08-361-920-29	Sequence 29, Appl
41	61	13.0	617	1	US-08-479-939-29	Sequence 29, Appl
42	61	13.0	617	1	US-08-483-432-29	Sequence 29, Appl
43	60.5	12.9	240	1	US-08-488-113B-147	Sequence 147, App
44	60.5	12.9	240	1	US-08-477-484B-147	Sequence 147, App
45	60.5	12.9	240	2	US-08-646-360-147	Sequence 147, App

ALIGNMENTS

RESULT 1
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.6%; Score 195.5; DB 4; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.2e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSSTLYVNC 60
: ||| ||| : : : ||| ||| ||| : ||| : |||
Db 1 MGSVKVLYKNNETSASTGSIKPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 59
: ||| ||| : : : ||| ||| ||| : ||| : |||

QY 61 WAAMGCNIRASFSGSVNPATPTADTYLQ 88
: ||| : : : ||| : : ||| :
Db 60 WAOTGASNVTFNFVKLSSGVSGADYYLE 87

RESULT 2

US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 41.4%; Score 194.5; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 5.7e-13;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSSTLYVNC 63
: ||| ||| : : : ||| ||| ||| : ||| : |||
Db 678 GVKVLYKNNETSASTGSIKPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 736
: ||| ||| : : : ||| ||| ||| : ||| : |||

QY 64 MGCNIRASFSGSVNPATPTADTYLQ 88
: ||| : : : ||| : : ||| :
Db 737 IGASNVTFNFVKLSSGVSGADYYLE 761

RESULT 3

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 40.7%; Score 191.5; DB 4; Length 1426;
Best Local Similarity 43.5%; Pred. No. 9.5e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSSSTLYVNC 63
: ||| ||| : : : ||| ||| ||| : ||| : |||
Db 413 GLKVLKNNETSASTGSIKPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 471
: ||| ||| : : : ||| ||| ||| : ||| : |||

QY 64 MGCNIRASFSGSVNPATPTADTYLQ 88
: ||| : : : ||| : : ||| :
Db 472 IGASNVTFNFVKLSSGVSGADYYLE 496


```

> TITLE OF INVENTION: In Stone Washing
> NUMBER OF SEQUENCES: 1
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
> STREET: 405 Lexington Avenue
> CITY: New York
> STATE: NY
> COUNTRY: U.S.A.
> ZIP: 10174
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Diskette
> COMPUTER: IBM Compatible
> OPERATING SYSTEM: DOS
> SOFTWARE: FastSeq for Windows Version 2.0
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/09/033,537A
> FILING DATE: 02-MAR-1998
> CLASSIFICATION: 008
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 0993/95
> FILING DATE: 08-SEP-1995
> APPLICATION NUMBER: PCT/DK96/00364
> FILING DATE: 03-SEP-1996
> ATTORNEY/AGENT INFORMATION:
> NAME: Green, Reza
> REGISTRATION NUMBER: 38,475
> REFERENCE/DOCKET NUMBER: 4492.204-US
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 212-867-0123
> TELEFAX: 212-878-9655
> TELEX:
> INFORMATION FOR SEQ ID NO: 1:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 551 amino acids
> TYPE: amino acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> US-09-033-537A-1

Query Match          35.2%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred.No. 2.2e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2

Qy      2 SGGVKVQYKNDSAPGNDITKGQLVLNQTSGSSVDLSTVVRYWFTRRDGGSTLVYNCDW 61
Db      402 TGNLVQYKVGDTSATDNQMKSFNKKNGTTPNLISGLKLRYTFKD-GTADMSASFDW 460
               :|::|||::||::|||::||::||::||::||::||::||::||::||::||::||::||
Qy      62 AAMCGCNRASFSGSVNPATPTADTYLQ 88
Db      461 AQIGASNVSAAF--ANFTGSNTDTYYE 485
               |::||::||::||::||::||::||::||::||::||::||::||::||

RESULT        6
PCT-US95-13813-9
Sequence 9, Application PC/TUS9513813
GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co. Ltd.
APPLICANT: Ramot University Authority for Applied
APPLICANT: Research and Industrial Development Ltd.
APPLICANT: Technion Research and Development Foundation Ltd.
APPLICANT: Bayer, Edward A.
APPLICANT: Morag, Ely
APPLICANT: Wilchek, Meir
APPLICANT: Lamed, Raphael
APPLICANT: Shoham, Yuval
TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
TITLE OF INVENTION: PROTEINS AND USE THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brody and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
```


STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13813
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.1%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 5.7e-11;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSPGDNOIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 5 VSGNLKVEFYNSPDTTNSINPQKVTNTGSSAIDLSKLTLYRYTVDCQKQDTFW-CD 63
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 64 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 100

RESULT 7
US-09-198-956-10
Sequence 10, Application US/09198956
Patent No. 6165769
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schulein, Martin
APPLICANT: Lange, Niels Erik K.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Schnorr, Kirk
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
FILE OF INVENTION: Licheniformis
FILE REFERENCE: 5377.200-US
CURRENT APPLICATION NUMBER: US/09/198,956
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: 1344/97
EARLIER FILING DATE: 1997-11-24
EARLIER APPLICATION NUMBER: 60/067,240
EARLIER FILING DATE: 1997-12-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 493
TYPE: PRT
ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSPGDNOIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 331 VSGNLKVEFYNSPDTTNSINPQKVTNTGSSAIDLSKLTLYRYTVDCQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 390 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 8
US-09-198-955A-12
Sequence 12, Application US/09198955A
Patent No. 6187580
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schulein, Martin
APPLICANT: Lange, Niels E.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Moller, Soren
APPLICANT: Glad, Sanne O. S.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars
TITLE OF INVENTION: NO. 6187580el Pectate Lyases
FILE REFERENCE: 5378.200-US
CURRENT APPLICATION NUMBER: US/09/198,955A
CURRENT FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: 1343/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 1344/97
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/067,249
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 60/067,240
PRIOR FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 09/073,684
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 09/184,217
PRIOR FILING DATE: 1998-11-02
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 493
TYPE: PRT
ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSPGDNOIKPGLQLVNTGSSVDLSTVTYVWFTRDGGSTLVNCD 60
DB 331 VSGNLKVEFYNSPDTTNSINPQKVTNTGSSAIDLSKLTLYRYTVDCQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
DB 390 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 9
US-07-862-588B-7
Sequence 7, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnaa
APPLICANT: Schulein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7


```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
;
; PRIOR APPLICATION DATA: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-07-862-588B-7

```

```

Query Match      25.6%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. NO. 1.9e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY  2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNCW 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  432 TGNLVQYKVGDISATDNQKPSFNKNGTTPVNLGLKXXXXKD-GPADMSCSIDW 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  62 AAMCGNIRASFSGVNPATPTADTY 86
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  491 AQIGRTNVLLAF--ANFTGNTDITY 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doh, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-048-164A-2

```

```

Query Match      24.3%; Score 114; DB 1; Length 162;
Best Local Similarity 29.9%; Pred. No. 2.3e-05;
Matches 29; Conservative 22; Mismatches 34; Indels 12; Gaps 4;

QY  2 SGGVKVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSTLVYNCW 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  3 TSSMSVEFYNSKSAQTSITPIKTIKNTSDSLNLDNVRVRYTSDGTGGQTFW-CDH 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY  62 AAMCGN-----IRASF--GSVNPATPTADTYLQ 88
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  62 AGALLGNSYVDNTSKVTANFVKETASP-TSTYDTYVE 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-08-460-462-2
; Sequence 2, Application US/08460462
; Patent No. 5670623
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doh, Roy H.
; TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,462
; FILING DATE: concurrently herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,164
; FILING DATE: 14-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE

```


Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	191	40.6	1711	2	T31337	1,4-beta-glucanase	
2	190.5	40.5	1779	2	T31085	xylanase - Caldice	
3	190	40.4	508	2	G69593	cellulase (EC 3.2.	
4	190	40.4	508	2	A26874	cellulase (EC 3.2.	
5	188	40.0	499	2	JN0111	cellulase (EC 3.2.	
6	186.5	39.7	1331	2	A48954	mannan endo-1,4-be	
7	186.5	39.7	1742	2	T17120	cellulase (EC 3.2.	
8	186	39.6	499	2	A27198	cellulase (EC 3.2.	
9	182.5	38.8	915	2	A43802	cellulase (EC 3.2.	
10	182.5	38.8	1039	2	S02711	cellulase (EC 3.2.	
11	180.5	38.4	145	2	A41897	cellulase homolog	
12	180	38.3	486	2	I40548	bifunctional cellu	
13	172	36.6	700	2	B41897	cellulase (EC 3.2.	
14	165.5	35.2	879	2	A47704	endoglucanase I (E	
15	165	35.1	1854	2	S36859	clpA protein - Clo	
16	149.5	31.8	505	2	S39962	cellulucanase - Er	
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2.	
18	141.5	30.1	986	2	S12021	thermoactive cellu	
19	130.5	27.8	586	2	PC6006	scaffolding protei	
20	130.5	27.8	1483	2	C37012	probably celluloso	
21	126.5	26.9	1162	2	T30433	scaffolding protei	
22	117	24.9	1230	2	S47466	cellulose 1,4-beta	
23	114	24.3	1848	2	A44140	cellulose-binding	
24	75.5	16.1	618	2	T08685	hypothetical prote	
25	73.5	15.6	547	2	T25478	hypothetical prote	
26	73.5	15.6	1428	2	AG2224	hypothetical prote	
27	70	14.9	5188	2	B85547	probable RTX fami	
28	70	14.9	5291	2	F69696	hypothetical prote	
29	68.5	14.6	574	2	A59196	cell surface glyco	

C;Genetics:


```

Db 362 GVSQVYRAGDGRVNSNQIRPOLHKNNGNATVDLKDVTARYWYVKNKGQN---FDCDIYA 418
      || |||| | |||| | : | : |||| || |||| : : | : : ||||:|
Qy 63 AMGCGNIRASFGSVNPATPTADTYLQ 88
      ||||| : | : : |||||
Db 419 QMCGCNLTHTKFTLHKPKQGADTYLE 444
      ||||| : | : : |||||

RESULT 5
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
C:Accession: JN0111
F:Park, S.H.; Kim, H.K.; Pack, M.Y.
Agric. Biol. Chem. 55, 441-448, 1991
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE616
A:Reference number: JN0111; MUID:91299280
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:D01057; NID:g216387; PIDN:BAA00859,1; PTD:d1001323; PID:g216388
A:Note: The authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 40.0%; Score 188; DB 2: Length 499;
Best Local Similarity 40.0%; Pred. No. 6,7e-12;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

Qy 4 GVKVQYKNNDSAPGDNQIKPOLQVLVNTGSSSVDSVLSTVTYRVYFTRDGSSTLVYVNCWAA 63
      | : |||| : : |||| | : | : |||| || |||| : ||||:|
Db 353 GISVQYRAGDGRVNSNQIRPOLHKNNGNATVDLKDVTARYWYVKNKGQN---DCDYAQ 410
      ||||| : | : : |||||

Qy 64 MCGGNIRASFGSVNPATPTADTYLQ 88
      ||||| : | : : |||||
Db 411 LCGGNVYKFTLHKPKQGADTYLE 435
      ||||| : | : : |||||

RESULT 6
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887,1; PID:g144291
A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIP:121577)
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPROHQHQRQ' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861,1; PID:g144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954

```


A: Experimental source: strain IFO3034
C: Function:
A: Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose
A: Pathway: cellulose degradation
C: Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F: 1-36/Domain: signal sequence #status predicted <SIG>

Query Match 39.6%; Score 186; DB 2; Length 499;
Best Local Similarity 40.7%; Pred. No. 1.1e-11;
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNNSAPGCDNQLKGLQLVNVTSSVDLSTVTYVWF-TRDGGSTLVYNCDDWA 62
| : ||||| ||||| : ||||| ||||| ||||| : : : |||||
Db 353 GISVQYKAGDGVNSNIPOLHIKKNNGATVDLKVTRYWYNAKKNQN---FDCDYA 409
| : ||||| ||||| : ||||| ||||| ||||| : : : |||||

QY 63 AMGCNIRASFGSVNPATPTADTYLQ 88
| : ||||| ||||| : ||||| ||||| : : : |||||
Db 410 QIGCGNLTHKFVTLHLKPKQGDATYLE 435
| : ||||| ||||| : ||||| ||||| : : : |||||

RESULT 9
A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum s
N: Alternate names: endo-1,4-beta-glucanase
C: Species: Caldocellum saccharotycum
C: Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C: Accession: A43802
R: Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A: Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile
A: Reference number: A43802; MUID:91136262
A: Accession: A43802
A: Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A: Molecule type: DNA
A: Residues: 1-915 <SAU>
A: Cross-references: EMBL:X13602
C: Function:
A: Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose
A: Pathway: cellulose degradation
C: Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C: Keywords: glycosidase; hydrolase; polysaccharide degradation
F: 20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SVY>

```

Query Match          38.8%; Score 182.5; DB 2; Length 915;
Best Local Similarity 41.9%; Pred. No. 4.8e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GGVKQVKYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTVVRYWFTRRDGGSGSTLVYNCDA 62
   | : | | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 349 GQIKVLYANKETNSTTIRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAAVS-DWA 407

Qy 63 AMCCGNIRASFGSVNPATPTADTYLQ 88
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 408 QIGASNVTFKRVKLSNVSSVSGADYYLE 433

RESULT 10
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C:Species: Caldocellum saccharolyticum
C:date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: S02711
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A:title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo
A:reference number: S02711; MUID:89098398
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1-1039 <SAU>

```

A:Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C:Genetics:
A:Gene: celB
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1039/Product: cellulase #status predicted <MAT>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 1039;
Best Local Similarity 41.9%; Pred. No. 5.6e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GGKVKVYKNNDSPGDNQIKPGLQLVWGTGSSVDLSVTWRYWFTRGDGGSTLVYNCDA 62
 | ||| : : : : : | : | : | : | : | : | : | : | : | : | : | :
Db 420 GQIKVLANKETNTTWTIRPLWKVANGSGSSIDLRSVTIRYTWTVDGRAQSAYS-DWA 478

Qy 63 AMGCGRINRASFSGSYNPATPTADTYLQ 88
 : | : | :
Db 479 QIGASNVTFKRVKLSSSVSGADYYLE 504

RESULT 11
A41897
cellulase homolog - Bacillus lautus (fragment)
C:Species: Bacillus lautus
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C:Accession: A41897; S27498
E:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Reference number: A41897; MUID:92276330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662
A:Experimental source: PL236
A>Note: sequence extracted from NCBI backbone (NCBITP:104604)

```

Query Match          38.4%; Score 180.5; DB 2; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.1e-11;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

Qy 7 VQYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTDRGGSSTLVYNCDDWAAMGC 66
   :||: | : ||||| : ||:||||: ||:||||: : ||: |
Db 1 LQYRAADTAADNQIKPSEIKNNGTSAVDLSTLKIRYFTKDGSAAVNGW-IDWAQLGG 59

Qy 67 GNIRASFGSVNPATPTADTYLQ 88
   ||: ||| | : |||:
Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 12
I40548
bifunctional cellulase precursor - Bacillus sp.
C;Species: Bacillus sp.
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C;Accession: I40548
J;Han, S.J.; Yoo, Y.J.; Kang, H.S.
J; Biol. Chem. 270, 26012-26019, 1995
A;Title: Characterization of a bifunctional cellulase and its structural gene: the ce
A;Reference number: I40548; MUID:96029707
A;Accession: I40548
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-486 <RES>
A;Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576

```


RESULT 14

A47704
endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A47704
R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.
J. Gen. Microbiol. 139, 307-316, 1993
A:Title: Gene sequence and properties of Cell, a family E endoglucanase from Clostridium
A:Reference number: A47704; MUID:93171873
A:Accession: A47704
A:Status: preliminary
A:Molecule type: DNA; protein
A:Residues: 1-879 <HAZ>
A:Cross-references: GB:I04735; NID:gl144807; PID:AAA20892.1; PID:gl144808
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)
C:Keywords: glycosidase; hydrolase

```

Query Match      35.1%; Score 165; DB 2; Length 1854;
Best Local Similarity 36.1%; Pred. No. 6.9e-09;
Matches: 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKVGQYKNDGAPGDNQIKPGLQLVNTGSSSSVDLSTVTVRVYFTRGDSSTLVNCD 60
    |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
365 VSGNLKVFEEYKNSPDITNINPQFKVNTGCSAIDLSKLTIRYYTVYDGGKDTFFW-CD 423
    |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```


Qy 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 424 HAAIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: July 2, 2002, 09:14:14
Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:18:03 ; Search time 28.05 Seconds
(without alignments)
122.853 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	190	40.4	499	GUN1_BACSU	P07983 bacillus su
2	190	40.4	499	GUN2_BACSU	P10475 bacillus su
3	188	40.0	499	GUN3_BACSU	P23549 bacillus su
4	186.5	39.7	1331	MANB_CALSA	P22533 caldocellum
5	186.5	39.7	1742	GUNA_CALSA	P22534 caldocellum
6	182.5	38.8	1039	GUNB_CALSA	P10474 c endogluca
7	180.5	38.4	145	YCEA_PAECLA	P29718 paenibacill
8	172	36.6	700	GUNA_PAECLA	P29719 paenibacill
9	165.5	35.2	879	GUNI_CLOTM	Q02934 clostridium
10	165	35.1	772	CIPB_CLOTM	Q01866 clostridium
11	165	35.1	1053	CIPA_CLOTM	Q06851 clostridium
12	149.5	31.8	444	GUNN_ERWCA	Q59394 erwinia car
13	149.5	31.8	505	GUNV_ERWCA	Q47096 erwinia car
14	147.5	31.4	504	GUNW_ERWCA	Q59395 erwinia car
15	147.5	31.4	914	GUX2_CLOSR	P50900 clostridium
16	141.5	30.1	986	GUNZ_CLOSR	P23659 clostridium
17	114	24.3	1848	CBPA_CLOCL	P38058 clostridium
18	67.5	14.4	118	HV39_MOUSE	P01809 mus musculu
19	67.5	14.4	215	FLAI_MYCTVA	P95316 methanococc
20	65	13.8	523	PUR8_MVCTU	P71553 m bifunctio
21	64	13.6	1571	ATC5_YEAST	P32660 saccharomyc
22	63.5	13.5	247	FLAI_THEVO	P57719 thermoplasm
23	63.5	13.5	465	TY3H_SCHMA	O17446 schistosoma
24	63	13.4	953	YA42_HUMAN	Q10969 homo sapien
25	63	13.4	1581	PPRB_HUMAN	O15648 h peroxisom
26	62.5	13.3	269	EXOK_RHIME	P33693 rhizobium m
27	62.5	13.3	802	PTNE_MOUSE	P29352 mus musculu
28	62	13.2	1210	ICEN_PSEFL	P09815 pseudomonas
29	61.5	13.1	252	HIS6_RHOSH	P50937 rhodobacter
30	61.5	13.1	316	MUCB_PSEAE	P38108 pseudomonas
31	61.5	13.1	328	HAIQ_MOUSE	P14428 mus musculu
32	61.5	13.1	368	HAIW_MOUSE	P03991 mus musculu
33	61.5	13.1	678	YIHO_ECOLI	P32138 escherichia

ALIGNMENTS

RESULT 1

ID	GUN1_BACSU	STANDARD;	PRT;	499 AA.
AC	P07983;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)			
DE	(Cellulase).			
GN	BGLC OR GLD.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DLG;			
RX	MEDLINE=87194581; PubMed=3106328;			
RA	Robson L.M., Chambliss G.H.;			
RT	"Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";			
RL	J. Bacteriol. 169:2017-2025(1987)			
CC	- - CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.			
CC	- - SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M16185; AAA22496.1; ALT_INIT.			
DR	PIR; A26874; A26874.			
DR	HSSP; O85465; 1A3H.			
DR	InterPro; IPR001956; CBD_3.			
DR	InterPro; IPR001547; Glyco_hydro_F5.			
DR	Pfam; PF00942; CBD_3; 1.			
DR	Pfam; PF00150; cellulase; 1.			
DR	ProDom; PD001947; CBD_3; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL 1 29			
FT	CHAIN 30 499			
FT	ENDOGLUCANASE.			
FT	ACT_SITE 169 169			
FT	ACT_SITE 257 257			
FT	NUCLEOPHILE (BY SIMILARITY).			
FT	DOMAIN 350 499			
FT	CELLULOSE-BINDING (BY SIMILARITY).			
SQ	SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;			

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 43.0%; Pred. No. 4.4e-14;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;


```

DR Subtilist; BG10437; bglc.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EA2 CRC64;

Query Match 40.4%; Score 190; DB 1; Length 499;
Best Local Similarity 40.7%; Pred.No. 4.4e-14;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps

QY 4 GVKYVKKNDSPAGDNQIKGLQLVNTGSSVDLSVTIVRYWF-TRDGSGSTLYNCDWA 62
   |:|::| : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 353 GISVQYRAGDGMNSNQIRPOLKKNNGTIVDLKVTRYWYKAKNKGN---FDCDYA 409

QY 63 AMGCNTRASFSGSVNPATPTADTYLQ 88
   ::||::| : ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 410 QIGCGNVTHRFVLTKPKQGADTYLE 435

RESULT 3
GUN3_BACSU STANDARD; PRT; 499 AA.
ID GUN3_BACSU
OC P23549;
DC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE DE Carboxymethyl-cellulase (CMCase) (Cellulase).
GN BGLC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BSE616;
RX MEDLINE=91299280; PubMed=1368694;
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616.";
RL Agric. Biol. Chem. 55:441-448(1991).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commec
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to licensedsib-sib.ch).
-----
CC EMBL; D01057; BAA00859.1; -.
CC PIR; JN0111; JN0111.
CC HSP; 085465; IA3H.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00150; cellulase; 1.
```



```
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55169 MW; 2E821E3D8BBACA04 CRC64;

Query Match 40.0%; Score 188; DB 1; Length 499;
Best Local Similarity 40.0%; Pred. No. 7 4e-14;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 4 GVKVQYKNDSDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTDGGSSLTLYNCDDAA 63
DB 353 GISVQYRAGDSMNSQIRPOLQIKNGNTTVDLKDVTARYWYNAKNGQNV--DCDYAQ 410

QY 64 MCGCNIRASFGSVNPTPTADTYLQ 88
DB 411 LGGCNVYKFEVTLHKPKQGADTYLE 435

RESULT 4
MANB_CALSA STANDARD; PRT; 1331 AA.
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
DE mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93119139; PubMed=1476429;
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme.";
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luethi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01257; AAA71887.1; -.
DR EMBL; M36063; AAA72861.1; -.
DR PIR; B43745; B43745.
DR PIR; A48954; A48954.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; GlyCo_hydro_F5.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Hydrolase; Glycosidase; Cellulose degradation; Signal;
KW Multifunctional enzyme.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.
FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).
FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).
FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).
FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
FT ACT_SITE 162 162 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 338 338 T -> P (IN REF. 2).
FT CONFLICT 340 346 TPTPTPT -> RQHQRQ (IN REF. 2).
SQ SEQUENCE 1331 AA; 146892 MW; PFBCA51BB8D8F0E0 CRC64;

Query Match 39.7%; Score 186.5; DB 1; Length 1331;
Best Local Similarity 42.5%; Pred. No. 3.5e-13;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNDSDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTDGGSSLTLYNCDDW 61
DB 364 SGOIKVLYANKETNTTIRPWLKVVNSGSSSIDLSRVTRYWTVVDGERAQSATS-DW 422

QY 62 AAMCGCNIRASFGSVNPTPTADTYLQ 88
DB 423 AQIGASNVTFKFKVLSSSVSGADTYLE 449

RESULT 5
GUNA_CALSA STANDARD; PRT; 1742 AA.
AC P22534;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'O V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RT extreme thermophile Caldocellum saccharolyticum.";
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).
RN [2]
RP SEQUENCE OF 1516-1742 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
```


QY 63 AMCGNIRASFGSVNPATPTADTYLQ 88
Db 479 QIGASNVTFKVLSSSVSGADYILE 504

RESULT 7
YCEA_PAELA STANDARD; PRT; 145 AA.
AC P29718;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in CELA 5' region (Fragment).
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Joergensen P.L., Diderichsen B.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase".
RL J. Bacteriol. 174:3522-3531(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76588; AAA22302.1; -
CC PIR: A41897; A41897.
CC PIR: S27498; S27498.
CC HSSP: Q06851; INBC.
CC InterPro: IPR001956; CBD_3.
CC Pfam: PF00942; CBD_3; 1.
CC ProDom: PD001947; CBD_3; 1.
CC Hypothetical protein.
KW NON_TER 1
FT SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;
SQ

Query Match 38.4%; Score 180.5; DB 1; Length 145;
Best Local Similarity 43.9%; Pred. No. 1.2e-13;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

QY 7 VOYKNNDSDPDNQIRKGLQVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCDDWAMGC 66
Db 1 LQYRAADTNAADNQIRKPSFNKNGTSAVDLSTLKIRYFTKDSAAVNGW-IDWAQLGG 59
QY 67 GNTIRASFGSVNPATPTADTYLQ 88
Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 8
GUNA_PAELA STANDARD; PRT; 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.

OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase".
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76588; AAA22303.1; -
CC PIR: B41897; B41897.
CC PIR: S27499; S27499.
CC HSSP: Q06851; INBC.
CC InterPro: IPR001956; CBD_3.
CC Pfam: PF00942; CBD_3; 1.
CC ProDom: PD001947; CBD_3; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700
FT ACT_SITE 213 213
FT SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EE0F CRC64;
SQ

Query Match 36.6%; Score 172; DB 1; Length 700;
Best Local Similarity 44.9%; Pred. No. 7.4e-12;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKNND-SAPGDNQIRKGLQVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNC 59
Db 549 VNSDLVVQYKDGDRNNATDNQIKPHFNQIKGTSPVDLSSLTRYFTKD-SSAAMNGWI 607
QY 60 DWAMCGGNIRASFGSVNPATPTADTYLQ 88
Db 608 DWAKLGSNIQISFGNHGA--DSDIYAE 634

RESULT 9
GUNI_CLOTM STANDARD; PRT; 879 AA.
AC Q02934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)
DE (Cellulase I).
GN CELI.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
RC STRAIN=NCIB 10682;
RX MEDLINE=93171873; PubMed=8436949;
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
RT Gilbert H.J.;
*Gene sequence and properties of CellI, a family E endoglucanase from

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L39788; AAC37033.1; -

CC HSSP; O85465; 1A3H.

CC InterPro; IPR001956; CBD_3.

CC InterPro; IPR001547; Glyco_hydro_F5.

CC Pfam; PF00942; CBD_3; 1.

CC Pfam; PF00150; cellulase; 1.

CC ProDom; PD001947; CBD_3; 1.

CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

CC Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT CHAIN 1 31 POTENTIAL.

FT DOMAIN 32 444 ENDOGLUCANASE N.

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.8%; Score 149.5; DB 1; Length 444;

Best Local Similarity 38.6%; Pred. No. 1.6e-09;

Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQYKNDGAPGNQIKPGQLVNTGSSVDLSTVTVRVWFTRDG--GSSLTVYNC 59

DB 357 TGDVVLQYRNVNDNPSDDAIRMAVNIKNTGSTPIKLSDLQVRYFFHDDGPGANLFV--- 413

QY 60 DWAMGCGNIRASFGSVNPATPTADTYL 87

DB 414 DWANVGNPNIVTGTGTPAASDTRANKRYV 441

RESULT 13

GUNW_ERWCA STANDARD; PRT; 505 AA.

AC Q47096;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V) (Cellulase V).

GN CELV.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCRI193;

RX MEDLINE=94067016; PubMed=8246888;

RA Cooper V.J.C., Saimond G.P.C.;

RT "Molecular analysis of the major cellulase (Celv) of Erwinia carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains.";

RT Mol. Gen. Genet. 241:341-350(1993).

RL -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X76000; CAA53592.1; -

DR HSSP; O85465; 1A3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_F5.

DR Pfam; PF00942; CBD_3; 1.

DR Pfam; PF00150; cellulase; 1.

DR ProDom; PD001947; CBD_3; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 335 352 LINKER.

FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.8%; Score 149.5; DB 1; Length 505;

Best Local Similarity 38.6%; Pred. No. 1.8e-09;

Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQYKNDGAPGNQIKPGQLVNTGSSVDLSTVTVRVWFTRDG--GSSLTVYNC 59

DB 354 TGDVVLQYRNVNDNPSDDAIRMAVNIKNTGSTPIKLSDLQVRYFFHDDGPGANLFV--- 410

QY 60 DWAMGCGNIRASFGSVNPATPTADTYL 87

DB 411 DWANVGNPNIVTGTGTPAASDTRANKRYV 438

RESULT 14

GUNW_ERWCA STANDARD; PRT; 504 AA.

AC Q59355;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1) (Cellulase V1).

GN CELV1.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCC3193;

RX MEDLINE=95231512; PubMed=7715600;

RA Mae A., Heikinheimo R., Palva E.T.;

RT "Structure and regulation of the Erwinia carotovora subspecies carotovora SCC3193 cellulase gene celv1 and the role of cellulase in photopathogenicity.";

RL Mol. Gen. Genet. 247:17-26(1995).

RL -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X79241; CAA55823.1; -

CC HSSP; O85465; 1A3H.

CC InterPro; IPR001956; CBD_3.

CC InterPro; IPR001547; Glyco_hydro_F5.

CC Pfam; PF00942; CBD_3; 1.

	Query Match	31.4%	Score 147.5	DB 1	Length 914
	Best Local Similarity	35.2%	Pred. No. 6.2e-09		
	Matches 31	Conservative 21	Mismatches 35	Indels 1	Gaps 1
Qy	1	VSGGVKVOYKNDSAPGDNQIKPGQLQLVNTGSSVLDSTVTYRVYWFTRDGGSSTLVY	NCD	60	
Db	763	VEGLYIIQSFENANTQEISNIMPRFINSYNGTSIPISEVKLYYYTVDSGDKPQNF	W-CD	821	
Qy	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88		
Db	822	WASIGSSNVGTGTFVKMDGATTGADYYLE	849		

Search completed: July 2, 2002, 09:18:04
Job time: 372 sec

RP CHARACTERIZATION.
RC STRAIN=NCIB 11754;
RX MEDLINE=91364686; Pubmed=1909625;
RA Bronnenmeier K., Ruecknagel K.P., Staudenbauer W.L.;
RT "Purification and properties of a novel type of
RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic
RT thermophile Clostridium stercorarium.";
RL Eur. J. Biochem. 200:379-385(1991).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL; 269359; CAA93280.1; -.
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001956; CBD 3.
DR	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:17:33 ; Search time 91.42 Seconds
(without alignments)
168.416 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKNNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	225.5	48.0	170	Q9REX6	Q9rfx6 caldibacill
2	216.5	46.1	930	Q9REX5	Q9rfx5 caldibacill
3	214.5	45.6	921	Q9L8L8	Q9l8l8 caldibacill
4	204.5	43.5	987	Q9Z4I1	Q9z4i1 bacillus sp
5	194.5	41.4	1751	Q9A0G4	Q9a0g4 caldicellul
6	192.5	41.0	1000	Q9A0G0	Q9a0g0 caldicellul
7	192.5	41.0	1770	Q9X3P5	Q9x3p5 caldicellul
8	191.5	40.7	261	Q9A0G7	Q9a0g7 caldicellul
9	191.5	40.7	1426	Q9X3P6	Q9x3p6 caldicellul
10	191	40.6	1711	Q963I1	Q963i1 anaerocellu
11	190.5	40.5	996	Q9AQH0	Q9aqh0 caldicellul
12	190.5	40.5	1779	Q52374	Q52374 caldicellul
13	190	40.4	499	Q93TJ6	Q93tj6 bacillus su
14	190	40.4	508	Q93LD0	Q93ld0 bacillus su
15	187	39.8	499	Q52731	Q52731 bacillus sp
16	186	39.6	499	Q45532	Q45532 bacillus su

17	186	39.6	501	2	083012	083012 bacillus sp
18	180	38.3	486	2	Q45430	Q45430 bacillus sp
19	172.5	36.7	1915	2	Q9REL0	Q9rpl0 acetivibrio
20	168	35.7	473	2	Q9RK75	Q9rk75 streptomyce
21	165.5	35.2	887	2	Q9L3J8	Q9l3j8 clostridium
22	150.5	32.0	2316	2	Q9FDJ9	Q9fdj9 bacteroides
23	130.5	27.8	1493	16	Q977Y4	Q977y4 clostridium
24	130.5	27.8	1546	2	Q45996	Q45996 clostridium
25	126.5	26.9	1162	2	Q82830	Q82830 clostridium
26	117	24.9	1230	2	Q59325	Q59325 clostridium
27	116.5	24.8	321	2	Q46392	Q46392 clostridium
28	87	18.5	440	2	Q93LI9	Q93li9 bacillus sp
29	78	16.6	135	13	Q90542	Q90542 ginglymosto
30	75.5	16.1	618	4	Q9Y3Z2	Q9y3z2 homo sapien
31	75.5	16.1	660	2	Q9L3J2	Q9l3j2 clostridium
32	75.5	16.1	986	4	Q9A858	Q9a858 homo sapien
33	75.5	16.1	1346	4	Q9UIZ3	Q9uiz3 homo sapien
34	73.5	15.6	547	5	P91006	P91006 caenorhabdi
35	73	15.5	511	5	Q9GYG5	Q9gyg5 caenorhabdi
36	72	15.3	499	12	Q993M3	Q993m3 autonomous
37	72	15.3	582	12	P90329	P90329 kilham rat
38	72	15.3	639	12	Q993M4	Q993m4 autonomous
39	72	15.3	722	12	P88901	P88901 kilham rat
40	71.5	15.2	263	17	Q9HID3	Q9hid3 thermoplas
41	71	15.1	389	3	Q9Y775	Q9y775 candida tro
42	70	14.9	137	13	Q9YHP9	Q9yhp9 ginglymosto
43	69.5	14.8	1430	16	Q98HK1	Q98hki rhizobium l
44	69	14.7	2873	12	Q93073	Q93073 hepatitis g
45	68.5	14.6	574	17	O26815	O26815 methanother

ALIGNMENTS

RESULT 1

Q9REX6 PRELIMINARY; PRT; 170 AA.
AC Q9REX6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN (FRAGMENT).
OS Caldicoccus cellulosivorans
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicoccus cellulosivorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22273.1; -;
DR HSSP; Q06851; INBC.
DR Interpro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.
DR ProDom; PD001947; CBD_3; 1.
FT Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F4E3A0B4 CRC64;

Query Match 48.0%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 2.6e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;
Qy 3 GGVKQYKNNDSAPGDNQIKPGLQVNTGSSSVSLVTYVWFTRDGGSTLVNCDWA 62
I : I I I I I : I I I I I : I I I I I : I I I I I : I I I I I :
Db 20 GSLVVQYRAADTNAGDNLKPKHFRIVNRTGSSVPLSELTIYRYTVD-GDKPQVFNCDWA 78
Qy 63 AMGCNTRASFGSVNPATPTADTYLQ 88

RT	"A novel thermostable multidomain 1,4-beta-xylanase from 'Caldibacillus cellulovorans' and effect of its xylan-binding domain on enzyme activity."									
RT	Microbiology 146:2947-2955(2000).									
RL	EMBL; AF200304; AAF61649.1; -									
DR	HSP; Q06851; INBC.									
DR	InterPro; IPR001956; CBD_3.									
DR	InterPro; IPR003305; CBD_6.									
DR	InterPro; IPR001000; Glyco_hydro_10.									
DR	InterPro; IPR002965; P-rich_extensn.									
DR	Pfam; PF02018; CBD_6; 1.									
DR	Pfam; PF00331; Glyco_hydro_10; 1.									
DR	PRINTS; PR00134; GLHYDRLASE10.									
DR	PRINTS; PR01217; PRICHEXTENS.									
DR	ProDom; PD001947; CBD_3; 2.									
KW	Signal; Xylan degradation; Hydrolase; Glycosidase.									
FT	SIGNAL 1 33 POTENTIAL.									
FT	CHAIN 34 921 BETA-1,4-XYLANASE XNNA.									
FT	SEQUENCE 921 AA; 102380 MW; C5DDDLIA7F7567413 CRC64;									
Query Match	45.6%; Score 214.5; DB 2; Length 921;									
Best Local Similarity	46.5%; Pred. No. 3.3e-15;									
Matches	40; Conservative 16; Mismatches 25; Indels 1; Gaps									
Qy	3 GGVKVVQKNNDSAPGQNIKPGILQVLNTGSSVDLSLTVRYWFTRDGSGSLVYNCDDWA 62									
Db	771 GNLVVQYRAADTNAGDNQLKPHRIVNRGTTSVPLSELTIRWYIVD-GDRPQVFNCDWA 829									
Qy	63 AMGCGNITRASFGSVNPATPTADYQLQ 88									
Db	830 WVGCSNLRGSLVLTGTGRTGADYYLE 855									
RESULT 4										
Q9Z4I1	PRELIMINARY; PRT; 997 AA.									
ID	Q9Z4I1									
AC	Q9Z4I1;									
DT	01-MAY-1999 (TrEMBLrel. 10, Created)									
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)									
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)									
DE	CELLULASE PRECURSOR (EC 3.2.1.4).									
CN	CEB.									
OS	Bacillus sp. BP-23.									
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;									
OC	Bacillus/Staphylococcus group; Bacillus.									
OX	NCBI_TaxID=89769;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
EC	STRAIN-BP-23;									
RX	MEDLINE=21129642; PubMed=11234960;									
RA	Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;									
RT	"Molecular cloning and characterization of a multidomain endoglucanase from Paenibacillus sp BP-23: evaluation of its performance in pulp refining";									
RT	Appl. Microbiol. Biotechnol. 55:61-68(2001).									
CC	-1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.									
DR	EMBL; AJ13614; CAB38941.1; -									
DR	HSP; P26221; 1TF4.									
DR	InterPro; IPR001956; CBD_3.									
DR	InterPro; IPR003961; FN.III.									
DR	InterPro; IPR001701; Glyco_hydro_9.									
DR	InterPro; IPR000566; Lipocin_cytFABP.									
DR	Pfam; PF00942; CBD_3; 2.									
DR	Pfam; PF00041; fn3; 2.									
DR	Pfam; PF00759; Glyco_hydro_9; 1.									
DR	ProDom; PD001947; CBD_3; 1.									
DR	SMART; SM00060; FN3; 2.									
DR	PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.									
DR	PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.									
DR	PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.									
KW	Glycosidase; Hydrolase; Signal.									
FT	SIGNAL 1 34 POTENTIAL.									

[illegible]


```

Qy 64 MGCNTRASFSGSVNPATPTADTYLQ 88
   1 1 1 : : : : : : : : : :
Db 903 IGASNVTFNFVKLSSGVSGADYYLE 927

RESULT 12
O52374 PRELIMINARY; PRT; 1779 AA.
ID AC O52374;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.8).
GN XYNL.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermocaneobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1.;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF036924; AAB95326.1; -
DR HSP; Q06851; INEC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00942; CBD_3; 3.
DR Pfam: PF02018; CBD_6; 2.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 40.5%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred.No. 3.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps

Qy 4 GVKYQKNDSAPGDNIKPLGLVNTGSSVDLSTVTVRYWFTDRGGSTLVNCDWAA 63
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1113 GLKVLKNNETSASTGSIRPFKIVNGSSVDLSRVKIRIYWTVDGDKPQSAV-CDWAQ 1171

Qy 64 MGCNTRASFSGSVNPATPTADTYLQ 88
   1 1 1 : : : : : : : : : :
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196

RESULT 13
Q93TU6 PRELIMINARY; PRT; 499 AA.
ID AC Q93TU6;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE ALKALI TOLERABLE CELLULASE.
GN CEL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y106;
RA Qi M., Wang P., Liu X., Qu Y.;
RT "The cloning and expression of a cellulase from Bacillus subtilis
RT Y106.";
```


RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF355629; AAK39540.1; -.
SQ SEQUENCE 499 AA; 55252 MW; 587A687584B7E5BE CRC64;

Query Match	40.4%	Score 190;	DB 2;	Length 499;
Best Local Similarity	40.7%;	pred. No. 8.1e-13;		
Matches	35;	Conservative	21;	Mismatches 26;
			Indels	4;
			Gaps	2;
Qy	4	GKVOYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTVRYWF--TRDGGSTLTVYNCOWA	62	
Db	353	GISVOYRAGDGSMSNQIRPOLQIKNGNTTVDLKDVTARYWYKAKNGQN--FDCDYA	409	
Qy	63	AMGCNTRASFGSVNPATPTADTYLQ	88	
Db	410	OIGCGNVTHKVTLHKPKOGADTVLE	435	

```

RESULT 14
Q93LDD0
ID Q93LDD0 PRELIMINARY; PRT; 508 AA.
AC Q93LDD0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHZ1;
RA Zvidzai C.J., Delgado O.D., Zvauya R., Mattiasson B.:
RT "Bacillus subtilis CHZ1 endo-b-1,4-glucanase gene.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY044252; AAK94871.1; -.
SQ HydroLase: Glycosidase.
KW SEQUENCE 508 AA; 56553 MW; A395EE1C6F3340F2 CRC64;

```

	Query Match	40.4%	Score 190;	DB 2;	Length 508;
	Best Local Similarity	40.7%;	Pred. No.	8.3e-13;	
	Matches	35;	Conservative	21;	Mismatches 26;
				Indels	4;
	Gaps				2;
Qy	4 GVKVOYKNNDSAPGDNOIKPGLQVLWNTGSSSDVLSVTVRVWF--TRDGGSSSTLYVNCDA 62				
	: :	: :	: :	: :	: :
Db	362 GISVOYRAGDCGSMNSQIRPOLQIKNNGTIVDLKDVTARYWYKAKKGQN---FCDYA 418				
Qy	63 AMGCGNIRASFGSVNPATPTADTYLQ 88				
	: :	: :	: :	: :	: :
Db	419 QIGCGNVTHKEVTLHKPKQGADTVLE 444				

RESULT	15	
ID	052731	PRELIMINARY; PRT; 499 AA.
AC	052731;	
DT	01-JUN-1998 (TrEMBLrel. 06, Created)	
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	ENDO-B-1,4-GLUCANASE (EC 3.2.1.4).	
CELS.		
OS	Bacillus sp. 79-23.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Bacillus.	
OX	NCBI_TaxID=72363;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	SPRAIN-79-23;	
RC	Yung K.H., Chun Y.C., Lee J.-C., Kim J.H., Yoon K.-H.;	
RL	"Cloning and expression of a Bacillus sp. 79-23 cellulase gene."	
RT	Biotechnol. Lett 18:1077-1082(1996).	

RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=79-23;	
RC	Yoon K.-H., Jung K.H.;	
RL	Submitted (JAN1998) to the EMBL/GenBank/DBJ databases.	
RA	EMBL: AF045482; AAC02536.1; -.	
DR	HSSP: O85465; IA3H.	
DR	DR	
DR	InterPro: IPR001956; CBD_3.	
DR	InterPro: IPR001547; Glyco_hydro_F5.	
DR	Pfam: PF00942; CBD_3; 1.	
DR	Pfam: PF00150; cellulase; 1.	
DR	Prodom: PD001947; CBD_3; 1.	
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.	
DR	Hydrolase; Glycosidase.	
SW	SEQUENCE 499 AA: 55193 MW: 990A1878788CF738 CRC64;	
SO		

```
Query Match      39.8%; Score 187; DB 2; Length 499;
Best Local Similarity 40.7%; Pred. No. 1.8e-12;
Matches 35; Conservative 20; Mismatches 27; Indels 4; Gaps 2;
```

```
QY    4   GVKVQVKNNDSAPGDNOIKGLQLVNTGSSVDLSTVTYRYWF--TRDGGSGSLTVYNCDWA 62
       |||: | :|||: |||: |::||| | |||: :: | : |||: |
Db    353 GISVQYRAGDGSMNSQIRPQLQIKNNGNTVDLKDVTARYWYKAKNGQNV---DCDYA 409

QY    63   AMCGGNIRASFGSVNPATPTADTYLQ 88
       :|||: | ::: ||||:
Db    410 QIGCGNVTHRFVTLHKPKQGADTYLE 435
```

Search completed: July 2, 2002, 09:17:34
Job time: 342 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:15:51 ; Search time 101.03 seconds
(without alignments)
96.748 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKVOYKKNDSAPGDNQ.....TRASPGSVNPATPTADTVLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	231.5	49.4	782	AA15625	Cellulase AE-1. A
2	195.5	41.7	616	AAV13494	Truncated cellulase
3	194.5	41.5	1751	AAV13493	Truncated cellulase
4	191.5	40.8	1426	AAV13492	Truncated cellulase
5	186	39.7	499	AA42122	NK-1 cellulase. B
6	172	36.7	700	AA13227	Novel endoglucanase
7	165.5	35.3	551	AAW18790	Corrected Bacillus
8	165	35.2	167	AAW95080	Cellulose binding
9	165	35.2	476	AAV54123	A mannanase-linker
10	165	35.2	493	AAV28850	Pectate lyase-linker
11	165	35.2	493	AAV43218	Pectate lyase CBD

12	165	35.2	531	18	AAW15238	Scaffoldin protein
13	165	35.2	1853	19	AAW43108	C. thermocellum ce
14	155.5	33.2	1352	22	AAG63962	Amino acid sequenc
15	150.5	32.1	1350	22	AAG63963	Amino acid sequenc
16	120.5	25.7	531	16	AAW01503	60 kD endoglucanas
17	120.5	25.7	532	12	AA13229	Endoglucanase enco
18	114	24.3	162	15	AA63634	Cellulose binding
19	114	24.3	162	20	AAW90077	C. cellulovorans C
20	114	24.3	163	22	AAE05745	Clostridium cellu
21	114	24.3	256	22	AA81128	Cl7E2 OSPA constru
22	114	24.3	328	22	AAE05749	Chimeric S peptide
23	114	24.3	341	22	AAE05747	Clostridium cellu
24	114	24.3	438	22	AAE05748	Clostridium cellu
25	113	24.1	190	22	AAE05746	Clostridium cellu
26	113	24.1	382	20	AAV39952	Gaussia luciferase
27	112	23.9	154	20	AAW90081	C. cellulovorans C
28	112	23.9	156	20	AAW90080	C. cellulovorans C
29	75.5	16.1	986	21	AAV40440	Human brain-derive
30	75.5	16.1	1346	22	AAU04567	Human G-protein co
31	75.5	16.1	1346	22	AAU04581	Human G-protein co
32	75.5	16.1	1371	22	AAU04570	Human G-protein co
33	75	16.0	1723	22	ABG24680	Novel human diagno
34	72.5	15.5	69	22	AAU17759	Novel human respir
35	70.5	15.0	2965	19	AAW56450	Fragment F1029 of
36	69	14.7	2970	19	AAW56445	Fragment HJ1916 o
37	68	14.5	2873	19	AAW56441	Nasturtium xyloglu
38	67.5	14.4	857	17	AAW5287	Sequence of the si
39	67	14.3	236	15	AAW45442	Anti-erbB2 scfv.
40	66	14.1	237	17	AAW4020	Single-chain anti-
41	66	14.1	237	18	AAW15185	Human p53 regulato
42	65.5	14.0	1566	20	AAV06309	Protein sequence o
43	64	13.6	117	20	AAW86130	Murine monoclonal
44	64	13.6	143	20	AAV50155	Drosophila melanog
45	64	13.6	254	22	ABB64503	

ALIGNMENTS

RESULT 1

AA15625

ID AA15625 standard; Protein; 782 AA.

AC AA15625;

XX 17-MAR-1992 (first entry)

DT Cellulase AE-1.

DE Detergents; pharmaceuticals; deinking; carboxymethylcellulose.

XX Aeromonas strain no. 212.

OS JP03251174-A.

PN 08-NOV-1991.

PD 28-FEB-1990; 90JP-0045465.

PF 28-FEB-1990; 90JP-0045465.

PR (OJIP) OJI PAPER KK.

PA WPI; 1991-373412/51.

XX N-PSDB; AAQ15178.

XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -

PT of opt. pH when carboxy-methyl-cellulose is used as substrate.

XX Claim 2; Fig 3; 8pp; Japanese.

XX The sequence was deduced from the gene whcih was sequenced from

CC plasmid, pABC 1, prepd. by ligating chromosomal DNA contg. the

CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt. of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-methylcellulose is the substrate. The N-terminal sequence: GIHADP- has been confirmed by Edman degradation. The gene can be used to produce recombinant enzyme which is used for the effective utilization of biomass resources and the mfr. of pharmaceuticals and foodstuffs, and also for the detergent and deinking of waste paper.

XX
SQ Sequence 782 AA;

Query Match 49.4%; Score 231.5; DB 12; Length 782;
Best Local Similarity 48.3%; Pred. No. 5e-18;
Matches 42; Conservative 16; Mismatches 28; Indels 1; Gaps 1;

QY 2 SGGVKVQYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVTVRVWFTRDGSSSLVYNCW 61
DQ 631 sgdllavdyktgdtnaadnqkphfnlknvkaavpilsrlryftad-gndqlqncdw 689
QY 62 AAMCGGNIRASFGSVNPATPTADTYLQ 88
DQ 690 amvgcslngafvkmnpkanadtyle 716

RESULT 2
AAY13494
ID AAY13494 standard; Protein; 616 AA.
XX
AC AAY13494;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulase Cel E3/B5.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI: 1999-315403/27.
DR N-PSDB; AAX55660.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 42-43; 65pp; English.
XX

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic

CC remove staining, backstaining or graying, for use on cellulosic materials including cotton-containing fabrics. They are especially useful for preventing redeposition of colorant during stonewashing, and for processing of textiles where cellulose breakdown is required. The new truncated enzymes show reduced redeposition of dye compared to using non-truncated cellulase compositions.

XX
SQ Sequence 616 AA;

Query Match 41.7%; Score 195.5; DB 20; Length 616;
Best Local Similarity 43.2%; Pred. No. 5.4e-14;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVKQYKNNDSAPGDNQIKPGQLVNTGSSVDLSTVTVRVWFTRDGSSSLVYNCW 60
DQ 1 mgsqgvklyknnetsastgsirpfwklvngsssvdlslrvkirywycvdgdkpqsav-cd 59
QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
DQ 60 waqigasvntfnvklssgvsgadyyle 87

RESULT 3
AAY13493
ID AAY13493 standard; Protein; 1751 AA.
XX
AC AAY13493;
XX
DT 30-JUL-1999 (first entry)
XX
DE Truncated cellulases comprising amino acid sequence.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing.
XX
OS Unidentified.
XX
PN EP921188-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-0810919.
XX
PR 19-SEP-1997; 97US-0932571.
XX
PA (CLRN) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
PI Gibbs MD, Morgan H, Williams DP;
XX
DR WPI: 1999-315403/27.
DR N-PSDB; AAX55662.
XX
PT New truncated cellulase proteins, useful in detergents and for
PT producing 'stonewashed' denim
XX
PS Claim 7; Page 37-41; 65pp; English.
XX

The invention relates to a recombinant cellulase active protein free of proteinases of native thermophilic and alkaliphilic origin, comprising the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3, Cel 6 or Cel E3/B5, or a stability region from one of the defined full-length sequences, or functional equivalents. Cel B5 extends from amino acid A1011 to P1424 or N1426, and Cel B4/5 extends from amino acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751 and the stability region extends from amino acid E482 to G635 in the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new enzymes are useful in laundry detergent compositions to prevent or remove staining, backstaining or graying, for use on cellulosic

Db 410 qicgcnthkfvltlhpkgqgadtyle 435

RESULT 6

AA13227
ID AAR13227 standard; Protein; 700 AA.

XX AC AAR13227;

XX DT 14-OCT-1991 (first entry)

XX DE Novel endoglucanase.

XX DE Cellulase activity; detergent.

XX KW Bacillus spp. NCIMB 40250.

XX OS

XX PH Key Location/Qualifiers

XX FT Peptide 1..31

XX FT /label= signal sequence

XX FT Protein 32..700

XX FT /label= mature endoglucanase

XX FT Cleavage-site 31..32

XX PN WO9110732-A.

XX XX 25-JUL-1991.

XX XX 18-JAN-1991; 91WO-DK00013.

XX XX 19-JAN-1990; 90DK-0000164.

XX XX (NOVO) NOVO NORDISK A/S.

XX XX Jorgensen PL, Schulein M, Hansen C;

XX XX WPI; 1991-238020/32.

XX XX N-PSDB; AAQ13001.

XX PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an

XX PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.

XX PT fabrics.

XX PS Claim 1; Page 80; 96pp; English.

XX CC The enzyme is produced by a strain of Bacillus spp. NCIMB 40250

XX CC and exhibits an endoglucanase activity of at least 10 (pref. at

XX CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total

XX CC protein under alkaline conditions. It is especially useful as a

XX CC cellulolytic agent and has been found to be more stable during

XX CC washing (60 mins. at 40 deg.) in the presence of conventional

XX CC detergents than a commercial cellulase preparation. It may also

XX CC show increased storage stability in liq. detergents contg.

XX CC proteases. The sequence was deduced from the DNA (AAQ13001).

XX CC See also AAR13228 and AAR13229.

XX SQ Sequence 700 AA;

Query Match 36.7%; Score 172; DB 12; Length 700;

Best Local Similarity 44.9%; Pred. No. 3.3e-11;

Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVLVNTGSSVDLSLTVTYRYWFTRDGSGSTLVYNC 59

Db 549 vnsdlvvykdgdrnatdqkphfnqkgtcpvdlssltiryftkd-ssaanngwi 607

QY 60 DWAMCGNIRASFGSVNPTPTADTYLQ 88

Db 608 dwaklgsgniqsfghnga--dsdtyae 634

RESULT 7

AAW18790

ID AAW18790 standard; protein; 551 AA.

XX AC AAW18790;

XX DT 18-NOV-1997 (first entry)

XX DE Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.

XX DE Endoglucanase; Endo 3A; formation; localised; variation;

XX KW colour density; surface; dye; fabric; family 5; cellulose;

XX KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;

XX KW blue jeans; back staining.

XX OS Bacillus lautus.

XX PN WO9709410-A1.

XX PD 13-MAR-1997.

XX PF 03-SEP-1996; 96WO-DK00364.

XX PR 08-SEP-1995; 95DK-0000993.

XX PA (NOVO) NOVO-NORDISK AS.

XX XX Fich M, Onishi M, Schulein M, Toft AH;

XX XX WPI; 1997-192888/17.

XX XX Localised variation of colour density in the surface of a dyed

XX PT cellulosic fabric - uses cellulase compsn. able to hydrolyse

XX PT p-nitrophenyl -beta-1,4-cellobioside

XX XX Disclosure; Pages 15-17; 23pp; English.

XX CC The present sequence is the corrected version of the incorrect

XX CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in

XX CC WO9110732. Endo 3 can be used in novel method of forming localised

XX CC colour density variation on the surface of a dyed cellulosic

XX CC fabric. The method comprises agitating the fabric in an aqueous

XX CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,

XX CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta

XX CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical

XX CC abrading agent or cellulose having abrading activity. Each

XX CC cellulase displays 30 % or more of its maximum activity at pH 7.

XX CC The process is useful to provide a stone washed look to blue jeans

XX CC without back staining.

XX SQ Sequence 551 AA;

Query Match 35.3%; Score 165.5; DB 18; Length 551;

Best Local Similarity 37.9%; Pred. No. 1.4e-10;

Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVOYKND-SAPGDNQIKPGLQVLVNTGSSVDLSLTVTYRYWFTRDGSGSTLVNCDW, 61

Db 402 tgnlvvykvvgtdsatdngmkpsfnknngtptpnlsiglkryyftkd-gtadmsasfdw 460

QY 62 AAMCGNIRASFGSVNPTPTADTYLQ 88

Db 461 aqigasnvsaaf--aftsntdtyve 485

RESULT 8

AA13227

ID AAR95080 standard; peptide; 167 AA.

XX AC AAR95080;

XX DT 30-OCT-1996 (first entry)

XX XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 09:12:55 ; Search time 39.87 Seconds
(without alignments)
53.912 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKVOYKNDSPGDNQ.....IRASFGSYNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.7	616	4	US-09-136-574A-47
2	194.5	41.5	1751	4	US-09-136-574A-44
3	191.5	40.8	1426	4	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	493	4	US-09-198-956-10
8	165	35.2	493	4	US-09-198-955A-12
9	120.5	25.7	531	2	US-07-862-588B-7
10	114	24.3	162	1	US-08-048-164A-2
11	114	24.3	162	1	US-08-460-462-2
12	114	24.3	162	1	US-08-460-457-2
13	114	24.3	162	1	US-08-460-458-2
14	114	24.3	162	2	US-08-460-455-2
15	114	24.3	162	3	US-08-330-394A-2
16	114	24.3	163	3	US-09-006-636-7
17	114	24.3	163	4	US-09-006-632-7
18	113	24.1	382	4	US-09-277-716-22
19	112	23.9	154	2	US-08-330-394A-29
20	112	23.9	156	2	US-08-330-394A-22
21	64	13.6	428	3	US-09-118-319-5
22	64	13.6	484	1	US-08-353-400-36
23	63.5	13.5	1785	4	US-09-341-587-3
24	63	13.4	1581	4	US-09-110-517-2
25	62.5	13.3	288	4	US-09-423-439-38
26	62.5	13.3	445	1	US-08-353-400-33
27	62.5	13.3	673	4	US-09-423-439-32

28	62.5	13.3	802	4	US-09-081-345-18	Sequence 18, Appl
29	62	13.2	1290	1	US-08-470-350B-2	Sequence 2, Appl
30	61.5	13.1	128	1	US-07-946-421-26	Sequence 26, Appl
31	61.5	13.1	307	2	US-08-484-905-68	Sequence 68, Appl
32	61.5	13.1	307	3	US-08-481-985B-68	Sequence 68, Appl
33	61.5	13.1	307	4	US-08-370-476-68	Sequence 68, Appl
34	61.5	13.1	316	1	US-08-260-202A-18	Sequence 18, Appl
35	61.5	13.1	316	1	US-08-017-114-18	Sequence 18, Appl
36	61.5	13.1	316	3	US-08-505-307-18	Sequence 18, Appl
37	61.5	13.1	316	5	PCT-US94-02034-18	Sequence 18, Appl
38	61	13.0	301	2	US-08-656-906-25	Sequence 25, Appl
39	61	13.0	301	4	US-09-217-847-25	Sequence 25, Appl
40	61	13.0	617	1	US-08-361-920-29	Sequence 29, Appl
41	61	13.0	617	1	US-08-479-939-29	Sequence 29, Appl
42	61	13.0	617	1	US-08-483-432-29	Sequence 29, Appl
43	60.5	12.9	240	1	US-08-488-113B-147	Sequence 147, App
44	60.5	12.9	240	1	US-08-477-484B-147	Sequence 147, App
45	60.5	12.9	240	2	US-08-646-360-147	Sequence 147, App

ALIGNMENTS

RESULT 1
US-09-136-574A-47
; Sequence 47, Application US/09136574A-
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 616 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 4; Length 616;
Best Local Similarity 43.2%; Pred. No. 1.2e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGVKVQYKNDSPAGDNOIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 60
Db 1 MGSQVLYKNNETSASTGSIKRWFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 59
QY 61 WAAGCGNIRASFGSVNPTPTADTYIQ 88
Db 60 WAQIGASNVTFNFVKLSGSGVADYILE 87

RESULT 2

US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA

COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 41.5%; Score 194.5; DB 4; Length 1751;
Best Local Similarity 44.7%; Pred. No. 5.7e-13;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNOIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 63
Db 678 GVKVLYKNNETSASTGSIKRWFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 736
QY 64 MGCNIRASFGSVNPTPTADTYIQ 88
Db 737 IGASNVTFNFVKLSGSGVADYILE 761

RESULT 3

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.

Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA

COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6294366e
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 40.8%; Score 191.5; DB 4; Length 1426;
Best Local Similarity 43.5%; Pred. No. 9.5e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNOIKPGLQVLVNTGSSSVLDLSTVTVRYWFTRDGSSSTLVYVNC 63
Db 413 GLKVLKNNETSASTGSIKRWFKIVNGSSSVLDLSRVKIRYWTVDGDKPQSAV-CD 471
QY 64 MGCNIRASFGSVNPTPTADTYIQ 88
Db 472 IGASNVTFNFVKLSGSGVADYILE 496

STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13813
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: BAYER-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 167 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match 35.2%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 5.7e-11;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLVNTGSSVDLSVTVRYWFTROGSGSTLVNCD 60
Db 5 VSNLKVVEFVNSPDTNSINPQFKVTNTGSSAIDLSKLTLYRYTVVDGQKQDTFW-CD 63
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 64 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 100

RESULT 7
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLVNTGSSVDLSVTVRYWFTROGSGSTLVNCD 60
Db 331 VSNLKVVEFVNSPDTNSINPQFKVTNTGSSAIDLSKLTLYRYTVVDGQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 8
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 2.1e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNNDSAPGDNOIKPGLQVLVNTGSSVDLSVTVRYWFTROGSGSTLVNCD 60
Db 331 VSNLKVVEFVNSPDTNSINPQFKVTNTGSSAIDLSKLTLYRYTVVDGQKQDTFW-CD 389
QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 9
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	191	40.7	1711	2	T31337	1,4-beta-glucanase
2	190.5	40.6	1779	2	T31085	xylianase (EC 3.2. cellulase
3	190	40.5	508	2	G69593	cellulase (EC 3.2. cellulase
4	190	40.5	508	2	A26874	cellulase (EC 3.2. cellulase
5	188	40.1	499	2	JN0111	cellulase (EC 3.2. mannan endo-1,4-be
6	186.5	39.8	1331	2	A48954	cellulase (EC 3.2. cellulase
7	186.5	39.8	1742	2	T17120	cellulase (EC 3.2. cellulase
8	186	39.7	499	2	A27198	cellulase (EC 3.2. cellulase
9	182.5	38.9	915	2	A43802	cellulase (EC 3.2. cellulase
10	182.5	38.9	1039	2	S02711	cellulase (EC 3.2. cellulase homolog
11	180.5	38.5	145	2	A41897	cellulase
12	180	38.4	486	2	I40548	bifunctional cellul
13	172	36.7	700	2	B41897	cellulase (EC 3.2. cellulase
14	165.5	35.3	879	2	A47704	endoglucanase I (E
15	165	35.2	1854	2	S36859	c1pA protein - Clo
16	149.5	31.9	505	2	S39962	endoglucanase - Er
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2. thermoactive cellu
18	141.5	30.2	986	2	S12021	scaffolding protei
19	130.5	27.8	586	2	PC6006	probably celluloso
20	130.5	27.8	1483	2	C97012	scaffolding protei
21	126.5	27.0	1162	2	T30433	cellulose 1,4-beta
22	117	24.9	1230	2	S74666	cellulose-binding
23	114	24.3	1848	2	A44140	hypothetical prote
24	75.5	16.1	618	2	T08685	hypothetical prote
25	73.5	15.7	547	2	T25478	hypothetical prote
26	73.5	15.7	1428	2	AC2224	probable RTX fami
27	70	14.9	5188	2	B85547	hypothetical prote
28	70	14.9	5291	2	F00696	hypothetical prote
29	68.5	14.6	574	2	A69196	cellulurface glyco

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 1.4e-11;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWFTRDGGSSLTLYNCDMAA 63

Db 1113 GLKLYLNNETSASTGSRPFKFIKVGSSVDLSRKYRYWTVDGDKPQSAV-CDWAQ 1171

QY 64 MCGGNIRASFSGVNPATPTADTYLQ 88

Db 1172 IGASNVTFNFVKLTSGVSGADTYLE 1196

RESULT 3

G69593 cellulase (EC 3.2.1.4) bglc precursor - Bacillus subtilis

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C;Accession: G69593; A26114; I40353; S24239; S49103; I39803

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toynoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033

A;Accession: G69593

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-508 <KUN>

A;Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CAB13696.1; PID:ell183471;

R;Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Sell

Nucleic Acids Res. 14, 9159-9170, 1986

A;Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.

A;Reference number: A26114; MUID:87066783

A;Accession: A26114

A;Molecule type: DNA

A;Residues: 10-508 <MAC>

A;Cross-references: GB:X04689; NID:g39823; PIDN:CAA28392.1; PID:g39824

A;Experimental source: strain PAP115

A;Note: part of this sequence, including the amino end of the mature form, was confirmed

R;Lindahl, V.; Ae, K.; Trousmo, A.

Antonie Van Leeuwenhoek 66, 327-332, 1994

A;Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C

A;Reference number: I40353; MUID:95225656

A;Accession: I40353

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 10-291, 'N', 293-508 <LIN1>

A;Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777

R;Lindahl, V.; Ae, K.

submitted to the EMBL Data Library, June 1992

A;Reference number: S24239

A;Accession: S24239

A;Molecule type: DNA

A;Residues: 10-291, 'N', 293-508 <LIN2>

A;Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777

A;Experimental source: strain CK-2

R;Wolf, M.; Geczi, A.; Borriass, R.

submitted to the EMBL Data Library, December 1993

A;Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons

A;Reference number: S49103

A;Accession: S49103

A;Molecule type: DNA

A;Residues: 10-508 <WOI>

A;Cross-references: EMBL:229076; NID:g509266; PIDN:CAA82317.1; PID:g509267

R;Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

Korean J. Microbiol. 24, 236-242, 1986

A;Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis

A;Reference number: I39803

A;Accession: I39803

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ',

A;Cross-references: GB:M38634; NID:g142657; PIDN:AAA22300.1; PID:g142658

A;Experimental source: strain ATCC 6633

C;Comment: The low molecular weight of the mature protein suggests carboxyl-terminal

C;Genetics:

A;Gene: bglc

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

C;Pathway: cellulose degradation

C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

F;1-38/Domain: signal sequence #status predicted <SIG>

F;39-508/Product: cellulase #status predicted <MAI>

Query Match 40.5%; Score 190; DB 2; Length 508;

Best Local Similarity 40.7%; Pred. No. 4.3e-12;

Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWF-TRDGGSSLTLYNCDMA 62

Db 362 GISVQYRAGDGSMNSQIRPOLQIKNGNTVDLKDVTARYWKAKNGQN---FDCDYA 418

QY 63 AMCGNIRASFSGVNPATPTADTYLQ 88

Db 419 QIGCGNVTHKFTVTLHKPKQGADTYLE 444

RESULT 4

A26874

cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Bacillus subtilis

C;Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999

A;Accession: A26874; B26874

R;Robson, L.M.; Chambliss, G.H.

J. Bacteriol. 169, 2017-2025, 1987

A;Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.

A;Reference number: A26874; MUID:87194581

A;Accession: A26874

A;Molecule type: DNA

A;Residues: 1-508 <ROBI>

A;Cross-references: GB:M16185; NID:gl43007; PIDN:AAA22496.1; PID:gl43008

A;Experimental source: strain DLG

A;Accession: B26874

A;Molecule type: protein

A;Residues: 39-53 <ROB2>

A;Experimental source: strain DLG

A;Note: the authors believe Met-1 and Met-2 may be alternate initiators

C;Comment: The low molecular weight of the mature protein suggests carboxyl-terminal

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A;Pathway: cellulose degradation

C;Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; po

F;1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.5%; Score 190; DB 2; Length 508;

Best Local Similarity 43.0%; Pred. No. 4.3e-12;

Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWF-TRDGGSSLTLYNCDMA 62

Db 362 GVSVOYRAGDGRVNSNIRPOLHFKNGNATVDLKDVTARTYWNKNGQN---FDCDYA 418
Query Match 39.8%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 419 QMGGCNLTFRFVTLHKPKOGADTYLE 444
RESULT 5
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
C:Accession: JN0111
R:Park, S.H.; Kim, H.K.; Pack, M.Y.
Agric. Biol. Chem. 55, 441-448, 1991
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE616
A:Reference number: JN0111; MUID:91299280
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:D01057; NID:g216387; PIDN:BAA00859.1; PID:dl001323; PID:g216388
A:Note: The authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal processing
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A:Pathway: cellulose degradation
A:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>
Query Match 40.1%; Score 188; DB 2; Length 499;
Best Local Similarity 40.0%; Pred. No. 6.7e-12;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;
Db 4 GVKVOYKKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWTRDGGSTLVYNCWAA 63
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 353 GISVOYRAGDGMNSNIRPOLHFKNGNATVDLKDVTARTYWNKNGQN---DCDYA 410
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 64 MCGCNIRASFGSVNPTADTYLQ 88
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 411 LGGCNVTKFVTLHKPKOGADTYLE 435
RESULT 6
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQHQR' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954

C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 39.8%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.8e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 2 SGGVKVOYKKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWTRDGGSTLVYNCW 61
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 364 SGQIKVLYANKETNTTIRPWLKVVNSGSSIDLSTRTYVWTRDGGSTLVYNCW 422
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 62 AAMGCGNIRASFGSVNPTADTYLQ 88
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 423 AQIGASNVTFKFKLSSVSGADTYLE 449
RESULT 7
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme thermophilic bacterium Caldocellum saccharolyticum
A:Reference number: Z18698; MUID:95336703
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:g537499; PID:g537500; PIDN:AAA91086.1
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cels
C:Keywords: glycosidase; hydrolase
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 2 SGGVKVOYKKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWTRDGGSTLVYNCW 61
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 704 SGQIKVLYANKETNTTIRPWLKVVNSGSSIDLSTRTYVWTRDGGSTLVYNCW 762
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 62 AAMGCGNIRASFGSVNPTADTYLQ 88
Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
Db 763 AQIGASNVTFKFKLSSVSGADTYLE 789
RESULT 8
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAK>
A:Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671

A;Experimental source: strain IFO3034

C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A;Pathway: cellulose degradation
C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F;1-36/Domain: signal sequence #status predicted <SIG>

Query Match 39.7% Score 186; DB 2; Length 499;
Best Local Similarity 40.7% Pred. No. 1.1e-11;
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

Qy 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 353 GISVQYKAGDGVNSQIRPOLHIKNKGNAFVLDKDVARTARYWYNAKNKQGN---FDCCDYA 409

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 410 QIGCNLTHKEVTLHKPKQGADTYLE 435

RESULT 9

A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum s
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Caldocellum saccharolyticum
C;Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C;Accession: A43802

R;Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990

A;Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "

A;Reference number: A43802; MUID:91136262

A;Accession: A43802
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A;Molecule type: DNA

A;Residues: 1-915 <SAU>

A;Cross-references: EMBL:X13602

C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce

A;Pathway: cellulose degradation

C;Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.9% Score 182.5; DB 2; Length 915;
Best Local Similarity 41.9% Pred. No. 4.8e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 349 GQIKVLYANKETNTTIRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAYS-DWA 407

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 408 QIGASNTTFKFKLVSSVSGADTYLE 433

RESULT 10

S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N;Alternate names: endo-1,4-beta-glucanase
N;Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C;Species: Caldocellum saccharolyticum
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C;Accession: S02711

R;Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989

A;Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for ex

A;Reference number: S02711; MUID:89098398

A;Accession: S02711

A;Molecule type: DNA

A;Residues: 1-1039 <SAU>

A;Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C;Genetics:
A;Gene: celB

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A;Pathway: cellulose degradation
C;Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-28/Domain: signal sequence #status predicted <SIG>
F;23-1039/Product: cellulase #status predicted <MAT>
F;72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.9% Score 182.5; DB 2; Length 1039;
Best Local Similarity 41.9% Pred. No. 5.1e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

Qy 3 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 62

Db 420 GQIKVLYANKETNTTIRPWLKVVNSGSSSIDLSRVTIRYWTVDGERAQSAYS-DWA 478

Qy 63 AMGCNTRASFSGSVNPATPTADTYLQ 88

Db 479 QIGASNTTFKFKLVSSVSGADTYLE 504

RESULT 11

A41897

cellulase homolog - Bacillus lautus (fragment)

C;Species: Bacillus lautus

C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999

C;Accession: A41897; S27498

R;Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A;Title: celA from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-

A;Reference number: A41897; MUID:92276330

A;Accession: A41897

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-145 <HAN>

A;Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662

A;Experimental source: PL236

A;Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 38.5% Score 180.5; DB 2; Length 145;
Best Local Similarity 43.9% Pred. No. 1.1e-11;
Matches 36; Conservative 18; Mismatches 25; Indels 3; Gaps 2;

Qy 7 VQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWF-TRDGGSTLVYNCDDWA 66

Db 1 LQYRAADTNAADNQIKPSFNKNGTSAVDLSTLKIRYFTKDGSAVNGW-IDWAQILGG 59

Qy 67 GNTRASFSGSVNPATPTADTYLQ 88

Db 60 SNIQISFG--NHTGTNSDTYVE 79

RESULT 12

I40548

bifunctional cellulase precursor - Bacillus sp.

C;Species: Bacillus sp.

C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999

C;Accession: I40548

R;Han, S.J.; Yoo, Y.J.; Kang, H.S.

J. Biol. Chem. 270, 26012-26019, 1995

A;Title: Characterization of a bifunctional cellulase and its structural gene: the ce

A;Reference number: I40548; MUID:96029707

A;Accession: I40548

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-486 <RES>

A;Cross-references: EMBL:U27084; NID:g857575; PIDN:AAC43478.1; PID:g857576

Qy 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
||:
Db 424 HAAIGSNGSYNGITSNVKGTFFVKMSSTNNADTYLE 460

Search completed: July 2, 2002, 09:14:14
Job time: 142 sec

GN	CELA.
OC	Paenibacillus lautus (Bacillus lautus).
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Paenibacillus.

OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN=ATCC 27405;
RX MEDLINE=93302508; PubMed=8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
RA Damain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipA) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein CipA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolov F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713(1997).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
CC COMPONENTS OF THE CELLULOSE.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08665; ; NOT_ANNOTATED_CDS.
CC EMBL; X67506; CAA47840.1; ;
CC PIR; S36859; S36859.
CC PDB; 1ANU; 23-JUL-97.
CC PDB; 1AOH; 08-JUL-98.
CC PDB; 1NBC; 26-SEP-97.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR002102; Cohesin.
CC InterPro; IPR002105; Dockerin_1.
CC Pfam; PF00942; CBD_3; 1.
CC Pfam; PF00963; Cohesin; 9.

DR Pfam; PF00404; Dockerin_1; 2.
DR ProDom; PD001947; CBD_3; 1.
DR STRAIN=PS00018; EF_HAND; UNKNOWN_1.
RX PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
KW 3D-structure.
FT SIGNAL. 1 28 CELLULOSONAL SCAFFOLDING PROTEIN A.
FT CHAIN 29 1853 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 364 522 9 X 150 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 54 1694 1-1.
FT REPEAT 29 182 1-2.
FT REPEAT 183 322 1-3.
FT REPEAT 560 704 1-4.
FT REPEAT 724 866 1-5.
FT REPEAT 889 1031 1-6.
FT REPEAT 1054 1196 1-7.
FT REPEAT 1219 1361 1-8.
FT REPEAT 1384 1526 1-9.
FT REPEAT 1548 1690 1-10.
FT DOMAIN 1791 1846 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 1791 1814 2-1.
FT REPEAT 1824 1847 2-2.
FT CONFLICT 1615 1615 A -> AA (IN REF. 1).
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;
Query Match 35.2%; Score 165; DB 1; Length 1853;
Best Local Similarity 36.1%; Pred. No. 1.5e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;
QY 1 VSGGVKVOYKNDSDAPGDNQIKGLQLVNTGSSVDLSTVTYRWFTRDGSSTLVYVNC 60
Db 365 VSGNLKVEFYNSPDTTNSINQFKVNTGSSAIDLSKLTLYIYIVDGKDTFW-CD 423
QY 61 WAAM-----CGGNIRASFGSVNPATPTADTYLQ 88
Db 424 HAALIGSGSYNGITSNVKCTFVKMSSSTNNADTYLE 460
RESULT 12
GUNN_ERWCA
ID GUNN_ERWCA STANDARD; PRT; 444 AA.
AC C59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATROSEPTICA FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multi-enzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L39788; AAC37033.1; -

DR HSP; O85465; IA3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_F5.

DR Pfam; PF00942; CBD_3; 1.

DR Pfam; PF00150; cellulase; 1.

DR ProDom; PD001947; CBD_3; 1.

DR ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 33 352 LINKER.

FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 444;

Best Local Similarity 38.6%; Pred. No. 1.6e-09;

Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQKNDSPAGNQIKPGQLVNTGSSVDLSTVTVRYWFTRDG--GSSTLYNC 59

DB 357 TGDVVLQYRNVDNPNPSDDAIRMAVNIKTGSTPIKLSDLQVRYFYHDDGKPGANLFV--- 413

QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87

DB 414 DWANVGPNIVTGTGTPAASTDKANRYV 441

RESULT 13

GUNW_ERWCA STANDARD; PRT; 505 AA.

ID GUNW_ERWCA

AC Q47096;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)

DE (Cellulase V).

GN CELV.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCRI193;

RX MEDLINE=94067016; PubMed=8245888;

RA Cooper V.J.C., Salmon G.P.C.;

RT "Molecular analysis of the major cellulase (CelV) of Erwinia carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains".

RT Mol. Gen. Genet. 241:341-350(1993).

CC -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X76000; CAA53592.1; -

DR HSP; O85465; IA3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_F5.

DR Pfam; PF00942; CBD_3; 1.

DR Pfam; PF00150; cellulase; 1.

DR ProDom; PD001947; CBD_3; 1.

DR ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 33 352 LINKER.

FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

Query Match 31.9%; Score 149.5; DB 1; Length 505;

Best Local Similarity 38.6%; Pred. No. 1.8e-09;

Matches 34; Conservative 19; Mismatches 30; Indels 5; Gaps 2;

QY 2 SGGVKVQKNDSPAGNQIKPGQLVNTGSSVDLSTVTVRYWFTRDG--GSSTLYNC 59

DB 354 TGDVVLQYRNVDNPNPSDDAIRMAVNIKTGSTPIKLSDLQVRYFYHDDGKPGANLFV--- 410

QY 60 DWAAMGCGNIRASFGSVNPATPTADTYL 87

DB 411 DWANVGPNIVTGTGTPAASTDKANRYV 438

RESULT 14

GUNW_ERWCA STANDARD; PRT; 504 AA.

ID GUNW_ERWCA

AC Q59355;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)

DE (Cellulase V1).

GN CELV1.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCC3193;

RX MEDLINE=95231512; PubMed=7715600;

RA Mae A., Heikinheimo R., Palva E.T.;

RT "Structure and regulation of the Erwinia carotovora subspecies carotovora SCC3193 cellulase gene celV1 and the role of cellulase in phytopathogenicity".

RT Mol. Gen. Genet. 247:17-26(1995).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X79241; CAA55823.1; -

DR HSP; O85465; IA3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_F5.

DR Pfam; PF00942; CBD_3; 1.

Search completed: July 2, 2002, 09:18:05
Job time: 373 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	225.5	48.1	170	2	Q9RFX6	Q9rfx6 caldbibacill
2	216.5	46.2	930	2	Q9RFX5	Q9rfx5 caldbibacill
3	214.5	45.7	921	2	Q9RL8L	Q9rl8l caldbibacill
4	204.5	43.6	997	2	Q9Z411	Q9z411 bacillus sp
5	194.5	41.5	1751	2	Q9AQG4	Q9aqg4 caldicellull
6	192.5	41.0	1000	2	O24820	O24820 thermophili
7	192.5	41.0	1770	2	Q9X3P5	Q9x3p5 caldicellull
8	191.5	40.8	261	2	Q9AQG7	Q9aqg7 caldicellull
9	191.5	40.8	1426	2	Q9X3P6	Q9x3p6 caldicellull
10	191	40.7	1711	2	P9G311	P9g311 anaerocellull
11	190.5	40.6	996	2	Q9AQH0	Q9aqh0 caldicellull
12	190.5	40.6	1779	2	O52374	O52374 caldicellull
13	190	40.5	499	2	Q93TJ6	Q93tj6 bacillus su
14	190	40.5	508	2	Q93LD0	Q93ld0 bacillus su
15	187	39.9	499	2	O52731	O52731 bacillus sp
16	186	39.7	499	2	Q45532	Q45532 bacillus su

DE	BETA-GLUCANASE.
OS	thermophilic anaerobe NA10.
OC	Bacteria.
OX	NCBI_TaxID=67756;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NA10;
RA	Miyake K., Machida Y., Hattori K., Iijima S.;
RT	"Characterization of a multi-domain cellulase from an extremely
RL	thermophilic anaerobe strain NA10.";
RE	Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AB008029; BAA22939.1;
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001589; Actinin_act_bind.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR001000; Glyco_hydro_10.
DR	InterPro; IPR001547; Glyco_hydro_F5.
DR	Pfam; PF00942; CBD_3; 1.
DR	Pfam; PF00150; cellulase; 1.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS; PR00134; GLHYDRLASE10.
DR	PRODOM; PD001947; CBD_3; 1.
DR	PROSITE; PS00019; ACTININ.1; UNKNOWN1.
DR	PROSITE; PS00591; GLYCOSYL_HVDROL_F10; 1.
DR	PROSITE; PS00659; GLYCOSYL_HVDROL_F5; UNKNOWN1.
SO	SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match	41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity	43.7%; Pred. No. 9.9e-13;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps	

Qy	2	SGGVKVOYKNNDASPGDNQIKPGQLQVNTGSSSVDLSTVTTRYWFTRDGGSTLVYNCDW	61
Db	372	SGQIKVLVANKETNSTTIRPWLVKVNTGSSSIDLSRVTIRYWTVDGDKQAAYS-DW	430
		: : : : :	

Qy	52	AAMCGGNIRASFGSVNPATPTADTYLQ	88
Db	431	AQIGASNVTFRFKVSLSSVSGADYYLE	457
		: : : : : :	

RESULT	7
Q9X3P5	ID Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC	Q9X3P5;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	XYNA.
GN	XYNA.
OS	Caldicellulosiruptor sp. Tok7B.1.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Thermoanaerobacter group; Caldicellulosiruptor.
OX	NCBI_TaxID=80339;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=TOK7B.1;
RX	MEDLINE=20171169; PubMed=10706665;
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.;
RA	Bergquist P.L.;
RT	"Multi-domain and multifunctional glycosyl hydrolases from the extreme
RL	thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL	Curr. Microbiol. 40:333-340(2000).
DR	EMBL; AF078737; AAD30363.1;
DR	HSSP; Q06851; INBC.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR003305; CBD_6.
DR	InterPro; IPR001000; Glyco_hydro_10.
DR	InterPro; IPR002965; Prich_extensn.
DR	Pfam; PF00942; CBD_3; 3.
DR	Pfam; PF02018; CBD_6; 2.
DR	Pfam; PF00331; Glyco_hydro_10; 1.
DR	PRINTS; PR00134; GLHYDRLASE10.


```

QY 64 MGCNINIRASFGSVNPATPTADTYLQ 88
   : : : : : : : : : : : : : :
Db 903 IGASNVTFNFVKLSSGVSGADYYLE 927

RESULT 12
O52374 PRELIMINARY; PRT; 1779 AA.
AC
DT 01-JUN-1998 (T+EMBLrel. 06, Created)
DT 01-JUN-1998 (T+EMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (T+EMBLrel. 19, Last annotation update)
DE FAMILY 10 XYLANASE (EC 3.2.1.8).
GN XYN.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RC STRAIN=RT69B.1;
RC Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036924; AAB95326.1; -.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF00942; CBD_3; 3.
DR Pfam; PF02018; CBD_6; 2.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR PROSITE; PS00591; GLYCOSYL-HYDROL_F10; 1.
DR Xylan degradation; Hydrolase; Glycosidase.
KW SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;
SQ

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.9%; Pred. No. 3.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSAPGDNQIKPGLQIVNTGSSVDLSLSTVTYRYWFTRDGGSTLVNCDWAA 63
   : : : : : : : : : : : : : : : : : : : :
Db 1113 GLKVLXNNETSASTGSIRPWFKIVNGSSVDLSLRKIRYWIYVDGKPSAV-CDWAAQ 1171

QY 64 MGCNINIRASFGSVNPATPTADTYLQ 88
   : : : : : : : : : : : : : :
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196

RESULT 13
Q93TJ6 PRELIMINARY; PRT; 499 AA.
AC
DT 01-DEC-2001 (T+EMBLrel. 19, Created)
DT 01-DEC-2001 (T+EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T+EMBLrel. 19, Last annotation update)
DE ALKALI TOLERABLE CELLULASE.
GN CEL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC STRAIN=Y106;
RC Qi M., Wang P., Liu X., Qu Y.;
RA "The cloning and expression of a cellulase from Bacillus subtilis
RT Rt y106.";
```


THIS PAGE BLANK (USPTO)